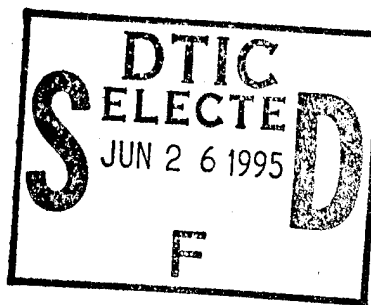


PL-TR-94-2296

## Evaluation of NWP and Cloud Forecasts From the Phillips Laboratory Global Spectral Model

Thomas Nehrkorn  
Mark Mickelson  
Marina Zivkovic  
Lawrence W. Knowlton



Atmospheric and Environmental Research, Inc.  
840 Memorial Drive,  
Cambridge, MA 02139

December 2, 1994

Final Report

May 1992 - November 1994

Approved for public release; distribution unlimited

19950623 018

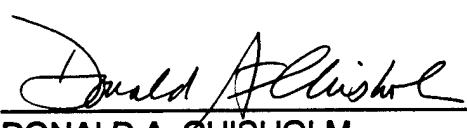



**PHILLIPS LABORATORY**  
**Directorate of Geophysics**  
**AIR FORCE MATERIEL COMMAND**  
**HANSCOM AIR FORCE BASE, MA 01731-3010**

DTIC QUALITY INSPECTED 8

"This technical report has been reviewed and is approved for publication."

  
DOUGLAS C. HAHN  
Contract Manager

  
DONALD A. CHISHOLM  
Chief, Satellite Analysis and Weather  
Prediction Branch  
Atmospheric Sciences Division

  
ROBERT A. McCLATCHEY, Director  
Atmospheric Sciences Division

This report has been reviewed by the ESC Public Affairs Office (PA) and is releasable to the National Technical Information Service (NTIS).

Qualified requestors may obtain additional copies from the Defense Technical Information Center (DTIC). All others should apply to the National Technical Information Service (NTIS).

If your address has changed, or if you wish to be removed from the mailing list, or if the addressee is no longer employed by your organization, please notify PL/IM, 29 Randolph Road, Hanscom AFB, MA 01731-3010. This will assist us in maintaining a current mailing list.

Do not return copies of this report unless contractual obligations or notices on a specific document requires that it be returned.

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.				
1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE December 2, 1994	3. REPORT TYPE AND DATES COVERED Final May 1992 - November 1994		
4. TITLE AND SUBTITLE Evaluation of NWP and Cloud Forecasts From the Phillips Laboratory Global Spectral Model		5. FUNDING NUMBERS F19628-92-C-0092 PE 62101F PR 4026 TA 01 WULD		
6. AUTHOR(S) Thomas Nehrkorn Marina Zivkovic Mark Mickelson Lawrence W. Knowlton				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Atmospheric and Environmental Research, Inc. 840 Memorial Drive Cambridge, MA 02139		8. PERFORMING ORGANIZATION REPORT NUMBER		
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Phillips Laboratory 29 Randolph Road Hanscom AFB MA 01731-3010 Contract Monitor: Douglas Hahn/GPAP		10. SPONSORING/MONITORING AGENCY REPORT NUMBER PL-TR-94-2296		
11. SUPPLEMENTARY NOTES				
12a. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited			12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 words)  This report describes the results of a 30-month effort aimed at evaluating forecasts of temperature, winds, humidity, and clouds from the advanced physics global spectral model (APGSM) developed by the Phillips Laboratory, in comparison with the currently operational GSM and cloud forecast models at Global Weather Central (GWC). Forecasts out to four days were generated for four separate months of 1989. Results show improved forecasts of thicknesses in the extratropics from the APGSM, but wind forecasts with generally larger errors. Forecasts of humidity are improved at some levels, regions, and forecast lead times, degraded at others. Root mean square errors of total cloud cover forecasts based on APGSM output are lower than those of the currently operational 5-layer model only for those techniques with much too smooth cloudiness distributions.				
14. SUBJECT TERMS Numerical weather prediction, Global spectral model, Cloud forecasting, TRNEPH data			15. NUMBER OF PAGES 226	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT SAR	

## Table of Contents

1.	Introduction	1
2.	Forecast Model Comparison	2
2.1.	The GWC GSM	2
2.2.	The APGSM	3
2.3.	Comparison Methodology	6
3.	Results of Forecast Model Comparison	8
3.1.	Results for January	8
3.2.	Results for Subsets of January Forecasts	11
3.3.	Results for April, July, and October	11
4.	Cloud Forecasts and Observations	12
4.1.	RTNEPH Data	12
4.2.	Regime-based Cloud Forecast Schemes	14
4.2.1.	Definition of weather regimes	14
4.2.2.	Average cloud cover schemes	20
4.2.3.	Cloud curve algorithm schemes	21
4.3.	Regression Schemes	25
4.3.	Slingo Scheme	33
4.4.	Persistence	33
4.5.	GWC 5-layer Forecasts	34
5.	Results of Cloud Forecast Verification	34
5.1.	Verification Statistics for January	36
5.2.	Verification Statistics for April, July, and October	37
5.3.	Comparison with GWC Verification Statistics	38
6.	Summary and Conclusions	40
7.	References	42
Appendix A:	Manuscript accepted for publication by the International Journal of Climatology	175
Appendix B:	Vertical stacking of layer cloud amounts	213

Accession For	
NTIS CRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification .....	
By .....	
Distribution / .....	
Availability Codes	
Dist	Avail and/or Special
A-1	

## **List of Figures**

1: Vertical distribution of the 15 analysis levels (HIRAS) and $\sigma$ -layers and interfaces for the GWC GSM (12-layers) and APGSM (18 layers).	4
2: Mean forecast error (bias) of 1000 hPa geopotential height for January forecasts, for the GWC GSM and APGSM. Errors are shown for the geographical regions of Table 1.	45
3: As Figure 2, except for the bias of 500 hPa geopotential height.	46
4: As Figure 2, except for the bias of 300 hPa geopotential height.	47
5: Vertical profile of mean errors of geopotential height for the APGSM (dotted line) and GWC GSM (solid line) for 0, 12, 24, ..., 96 hour forecasts, for the geographical regions of Table 1.	48
6: As Figure 2, except for the root mean square (RMS) error of 1000 hPa geopotential height.	49
7: As Figure 2, except for the RMS error of 500 hPa geopotential height.	50
8: As Figure 2, except for the RMS error of 300 hPa geopotential height.	51
9: As Figure 5, except for the RMS error of geopotential height.	52
10: As Figure 2, except for the anomaly correlation of 1000 hPa geopotential height.	53
11: As Figure 2, except for the climatological skill score of 1000 hPa geopotential height.	54
12: As Figure 2, except for the anomaly correlation of 500 hPa geopotential height.	55
13: As Figure 2, except for the climatological skill score of 500 hPa geopotential height.	56
14: As Figure 2, except for the RMS error of the horizontal wind at 850 hPa.	57

15: As Figure 2, except for the RMS error of the horizontal wind at 300 hPa.	58
16: As Figure 5, except for the RMS errors of the horizontal wind.	59
17: As Figure 5, except for the RMS errors of the zonal wind.	60
18: As Figure 5, except for the RMS errors of the meridional wind.	61
19: As Figure 2, except for the bias of 850 hPa RH.	62
20: As Figure 2, except for the bias of 500 hPa RH.	63
21: As Figure 2, except for the bias of 300 hPa RH.	64
22: As Figure 5, except for the bias of RH.	65
23: As Figure 2, except for the RMS error of 850 hPa RH.	66
24: As Figure 2, except for the RMS error of 500 hPa RH.	67
25: As Figure 2, except for the RMS error of 300 hPa RH.	68
26: As Figure 5, except for the RMS error of RH.	69
27: As Figure 5, except for the bias of geopotential for April.	70
28: As Figure 5, except for the bias of geopotential for July.	71
29: As Figure 5, except for the bias of geopotential for October.	72
30: As Figure 2, except for the RMS error of 1000 hPa geopotential height for April.	73
31: As Figure 2, except for the RMS error of 1000 hPa geopotential height for July.	74
32: As Figure 2, except for the RMS error of 1000 hPa geopotential height for October.	75
33: As Figure 2, except for the RMS error of 500 hPa geopotential height for April.	76

34: As Figure 2, except for the RMS error of 500 hPa geopotential height for July.	77
35: As Figure 2, except for the RMS error of 500 hPa geopotential height for October.	78
36: As Figure 5, except for the RMS errors of geopotential for April.	79
37: As Figure 5, except for the RMS errors of geopotential for July.	80
38: As Figure 5, except for the RMS errors of geopotential for October.	81
39: As Figure 5, except for the RMS errors of horizontal wind for April.	82
40: As Figure 5, except for the RMS errors of horizontal wind for July.	83
41: As Figure 5, except for the RMS errors of horizontal wind for October.	84
42: Mean profiles of the 24 original synoptic regimes. Shown are the temperature, wind speed, and RH, as a function of model sigma layer (going from 1 near the surface to 12 at $\sigma=.325$ (viz. Figure 1).	85
43: Sample sizes and contingency table scores of the combinations and individual regimes, for forecast lead times from 12 hours to 96 hours. Notice the change in scale for regime 23*.	91
44: The geographic distribution of the regimes together with contours of the 1000hPa - 500 hPa thickness field for 00 UTC 5 April 1989.	100
45: The geographic distribution of the regimes together with contours of the 300 hPa wind speed (every 10 m/s, only contours above 30 m/s shown) for 00 UTC 5 April 1989.	101
46: Histograms of total cloud cover for a one-week period in April for the 15 regimes.	102
47: CCA curves for 12 hour forecasts, using the average RH of the cloud layers, for the 15 regimes.	105
48: CCA curves for 24 hour forecasts, using the average RH of the cloud layers, for the 15 regimes.	108

49: The Northern Hemisphere octagon, shown in light shading over the RTNEPH grid. Axis labels are the distance from the lower left corner of the grid in units of 1/2-mesh grid lengths. Also shown in darker shading are areas encompassing the Eurasian and North American continents, denoted in the verification statistics as "backhalf" and "fronthalf", respectively.	111
50: Bias of 12-hour forecasts of cloud cover (Total and the six RTNEPH layers) of the candidate cloud cover schemes for the dependent data set. Slingo scheme is labeled sli, average cloud cover scheme with (without) overlap computation for total cloud cover is labeled m13 (m11), CCA scheme using average RH and separate total cloudcover curve (overlap computation) is labeled c11 (c13), corresponding maximum RH CCA schemes are labeled c22 and c23.	112
51: As Figure 50, except for the sharpness of 12-hour cloud cover forecasts. Sharpness of the verifying RTNEPH analyses is indicated by horizontal line.	114
52: As Figure 50, except for the RMS error of 12-hour cloud cover forecasts.	116
53: As Figure 50, except for the correlation between forecast (12 hour) and verifying cloud cover.	118
54: Scatterplot of 12-hour total cloud cover RMS errors of average RH CCA schemes. Scores for the scheme with separate total cloud cover CCA curves (c11) shown as function of the scores for the scheme with overlap computation (c13).	120
55: As Figure 50, except for the bias of 96-hour cloud cover forecasts.	121
56: As Figure 50, except for the RMS errors of 96-hour cloud cover forecasts.	123
57: As Figure 50, except for the correlation between forecast (96-hour) and verifying cloud cover.	125
58: Bias of 12-96 hour forecasts of cloud cover (Total and the six RTNEPH layers) of the 6 cloud cover schemes for January, computed over the NH RTNEPH octagon. Refer to text for explanation of curve labels.	127
59: Same as Figure 58, except for the bias of 12-96 hour forecasts of cloud cover, computed over the "fronthalf" portion of the NH RTNEPH octagon.	129



60: Same as Figure 58, except for the bias of 12-96 hour forecasts of cloud cover, computed over the "backhalf" portion of the NH RTNEPH octagon.	131
61: Same as Figure 58, except for the rmse of 12-96 hour forecasts of cloud cover, computed over the NH RTNEPH octagon.	133
62: Same as Figure 58, except for the rmse of 12-96 hour forecasts of cloud cover, computed over the "fronthalf" portion of the NH RTNEPH octagon.	135
63: Same as Figure 58, except for the rmse of 12-96 hour forecasts of cloud cover, computed over the "backhalf" portion of the NH RTNEPH octagon.	137
F64: Same as Figure 58, except for the correlation of 12-96 hour forecasts of cloud cover, computed over the NH RTNEPH octagon.	139
65: Same as Figure 58, except for the correlation of 12-96 hour forecasts of cloud cover, computed over the "fronthalf" portion of the NH RTNEPH octagon.	141
66: Same as Figure 58, except for the correlation of 12-96 hour forecasts of cloud cover, computed over the "backhalf" portion of the NH RTNEPH octagon.	143
67: Same as Figure 58, except for the sharpness of 12-96 hour forecasts of cloud cover, computed over the NH RTNEPH octagon.	145
68: Same as Figure 58, except for the sharpness of 12-96 hour forecasts of cloud cover, computed over the "fronthalf" portion of the NH RTNEPH octagon.	147
69: Same as Figure 58, except for the sharpness of 12-96 hour forecasts of cloud cover, computed over the "backhalf" portion of the NH RTNEPH octagon.	149
70: Same as Figure 58, except for the bias of 12-96 hour forecasts of cloud cover for April, computed over the NH RTNEPH octagon.	153
71: Same as Figure 58, except for the bias of 12-96 hour forecasts of cloud cover for July, computed over the NH RTNEPH octagon.	155
72: Same as Figure 58, except for the bias of 12-96 hour forecasts of cloud cover for October, computed over the NH RTNEPH octagon.	157

73: Same as Figure 58, except for the rmse of 12-96 hour forecasts of cloud cover for April, computed over the NH RTNEPH octagon.	159
74: Same as Figure 58, except for the rmse of 12-96 hour forecasts of cloud cover for July, computed over the NH RTNEPH octagon.	157
75: Same as Figure 58, except for the rmse of 12-96 hour forecasts of cloud cover for October, computed over the NH RTNEPH octagon.	161
76: Same as Figure 58, except for the correlation of 12-96 hour forecasts of cloud cover for April, computed over the NH RTNEPH octagon.	163
77: Same as Figure 58, except for the correlation of 12-96 hour forecasts of cloud cover for July, computed over the NH RTNEPH octagon.	165
78: Same as Figure 58, except for the correlation of 12-96 hour forecasts of cloud cover for October, computed over the NH RTNEPH octagon.	167
79: Same as Figure 58, except for the sharpness of 12-96 hour forecasts of cloud cover for April, computed over the NH RTNEPH octagon.	169
80: Same as Figure 58, except for the sharpness of 12-96 hour forecasts of cloud cover for July, computed over the NH RTNEPH octagon.	171
81: Same as Figure 58, except for the sharpness of 12-96 hour forecasts of cloud cover for October, computed over the NH RTNEPH octagon.	173

### **List of Tables**

1. List of geographical regions used for the verification of forecasts	6
2. Number of elements in retained clusters obtained in two step clustering procedure.	15
3. Numbering of weather regimes after clustering steps 1 and 2..	16

4: List of the final 15 weather regimes. First column is the regime number ("*"s indicate combined regimes). Second column lists the regime numbers from the original 24 regimes included in the combined regimes.	20
5: Average cloud cover for the regimes, for the six RTNEPH layers, and for total cloud cover from the RTNEPH data ( $T_r$ ) and from the overlap computation ( $T_o$ ).	21
6: PL scheme predictors from the forecast valid 6 hours before the cloud forecast valid time. The column labeled No contains the designation used by Norquist et al. (1994). Unless otherwise noted, all quantities are averages of the predictand deck. Adapted from Norquist et al. (1994).	27
7: Predictors valid at the verifying time used in the PL scheme. Unless otherwise noted, deck averages are used. Column labeled Deck indicates whether quantities are for all 3 decks (3), the predictand deck only (1), surface (S) or column-integrated (C). A "Y" entry in the column labeled Res indicates predictors used in the regression of the residuals from the average cloud cover scheme. Adapted from Norquist et al. (1994).	28
8: List of most frequently selected predictors for PL scheme. The first column refers to No given in Tables 6 and 7; second column gives the percentage a given predictor was selected; the last three columns are the average, minimum, and maximum step number of the stepwise regression when it was included	32
9: List of most frequently selected predictors for regression of residuals of average cloud cover scheme. See Table 8 for explanation.	33
10: Comparison of 24 hour total cloud cover rmse scores computed over the NH octagon. For each method, values are shown for the forecast rmse (Fcst), persistence rmse (Per), and the percentage improvement over persistence (Skill).	39
11: Evolution of GWC 5-layer total cloud cover correlation and rmse scores with forecast lead time (12-48 hours), for January, April, July, and October 1989, for the fronthalf (-f) and backhalf (-b).	40

## **1. Introduction**

This report describes the results of a 30-month study designed to demonstrate that the advanced weather prediction model developed at the Geophysics Directorate of the Phillips Laboratory (PL) is capable of fulfilling the forecasting needs of the US Air Force. Air Force operations are influenced in a significant way by the weather, and there is a need for accurate forecasts of not only the conventional forecast variables (pressure, temperature, moisture, and winds), but also other weather related variables such as aerosols, haze and clouds. Forecasts of these non-conventional variables must be based on sophisticated models that simulate the physical processes involved.

The Air Force Global Weather Central (GWC) currently uses a global spectral model (GSM) obtained from the National Meteorological Center (NMC) as far back as 1984, with a very simple set of physical parameterizations. Operational cloud forecasts are based on trajectory models with highly parameterized physics, and utilize only the wind fields of the forecast model output. The Phillips Laboratory has developed a replacement for this GWC GSM with advanced physics parameterizations (APGSM), and demonstrated its potential usefulness in preliminary tests. Simultaneously, cloud forecast schemes based on the forecast model output have been developed and tested as a potential replacement or complement to the present trajectory models.

In the present study, forecasts of conventional variables from both models (GWC GSM and APGSM) are compared in side-by-side tests, using data from all four seasons of 1989. The operational GWC analyses (High Resolution Analysis System, or HIRAS) are used for initializing and verifying the forecast models. The experimental design and results from this evaluation are described in Sections 2 and 3. The APGSM model output was used to generate cloud forecasts, which were verified against the operational GWC cloud analyses (RTNEPH). The cloud forecasts and observations are described in Section 4, and verification results are presented in Section 5. A summary and conclusions form the final section.

## **2. Forecast Model Comparison**

### **2.1. The GWC GSM**

The GSM currently operational at GWC is based on the model developed at NMC (Sela, 1980). The hydrodynamics were completely redesigned (Brenner et al., 1982). The physics package was last updated in 1984, and it consists of a very simple set of physical parameterizations. It does not simulate the transfer of radiation through the atmosphere, and includes only a drag-law type boundary layer parameterization. The adjustment physics consists of large-scale precipitation when model cells become saturated, dry adiabatic adjustment to avoid instability, and a version of the Kuo (1965) convection scheme, which is disabled at most points through the use of high threshold values and various criteria that must be met before moist convection is allowed to take place. This physics package is described more fully in Yang et al. (1989).

Model initial states are derived from gridded analyses by preprocessing (vertical interpolation to the model's  $\sigma$ -surfaces, and spectral transformation from gridpoint to spectral space), and a subsequent nonlinear normal mode initialization (NMI). The NMI is a standard adiabatic Machenhauer (1977) scheme obtained from NMC (Ballish, 1980).

In the present study, parameters of the GSM are used that closely correspond to current operational practice at GWC. In particular, the horizontal resolution is set at rhomboidal truncation 40, and 12 layers are used in the vertical (moisture is carried at only 7 layers). The sigma layer interfaces for the 12 layers are at 1.0, .925, .800, .650, .500, .375, .300, .250, .200, .150, .100, .050, 0. A time step of 12 minutes is used, and the coefficients for the time stepping are 0.04 (for the time filter coefficient) and 0.5 (for the time stepping coefficient, implying a semi-implicit step). The coefficient for the horizontal  $\nabla^4$  diffusion is  $6 \times 10^{15} \text{ m}^4 \text{ s}^{-1}$ . The terrain data set is a mean terrain height derived from a coarse input gridded field ( $2.5^\circ \times 2.5^\circ$ ), and the boundary data sets of sea surface temperature and drag coefficients use monthly mean values from NMC datasets. The NMI uses two iterations of the Machenhauer scheme, initializing only the four gravest modes with periods less than 48 hours.

The HIRAS analyses used for initialization and verification consist of geopotential height, zonal and meridional wind on 15 mandatory pressure levels

(1000 , 850 , 700 , 500 , 400 , 300, 250, 200, 150, 100, 70, 50, 30, 20, and 10 hPa), and relative humidity on the lowest 6 mandatory levels. Figure 1 shows the distribution of the analysis levels in the vertical, alongside those of the GWC GSM. It can be seen that while the resolution of the GSM is slightly higher than that of the analysis in the lower troposphere, it is much coarser in the stratosphere (the top 4 analysis levels are all contained in a single, the topmost,  $\sigma$ -layer).

## **2.2. The APGSM**

The APGSM hydrodynamics code is based on the code described by Brenner et al. (1982), but it was recoded by Nehrkorn et al. (1992) to allow more general horizontal truncation, and to make use of vectorization and multiprocessing. The physics package of the APGSM was developed by several research groups and tested and integrated by PL personnel. The version of the GSM used in this study is described in Norquist et al. (1992). It contains a planetary boundary layer (PBL) parameterization (Mahrt et al., 1984), which includes two soil layers and makes use of geographic databases of surface roughness, soil type, albedo, and other surface fields. A gravity wave drag parameterization (Vernekar et al., 1991) is included, as is a radiative transfer package developed by Liou et al. (1984), Ou et al. (1988), and Schattell (1992). The dry adiabatic adjustment and the large scale precipitation parameterization remain essentially unchanged from the GWC GSM, but moist convection is parameterized with the European Centre for Medium-Range Weather Forecasts (ECMWF) mass flux scheme (Tiedtke, 1989).

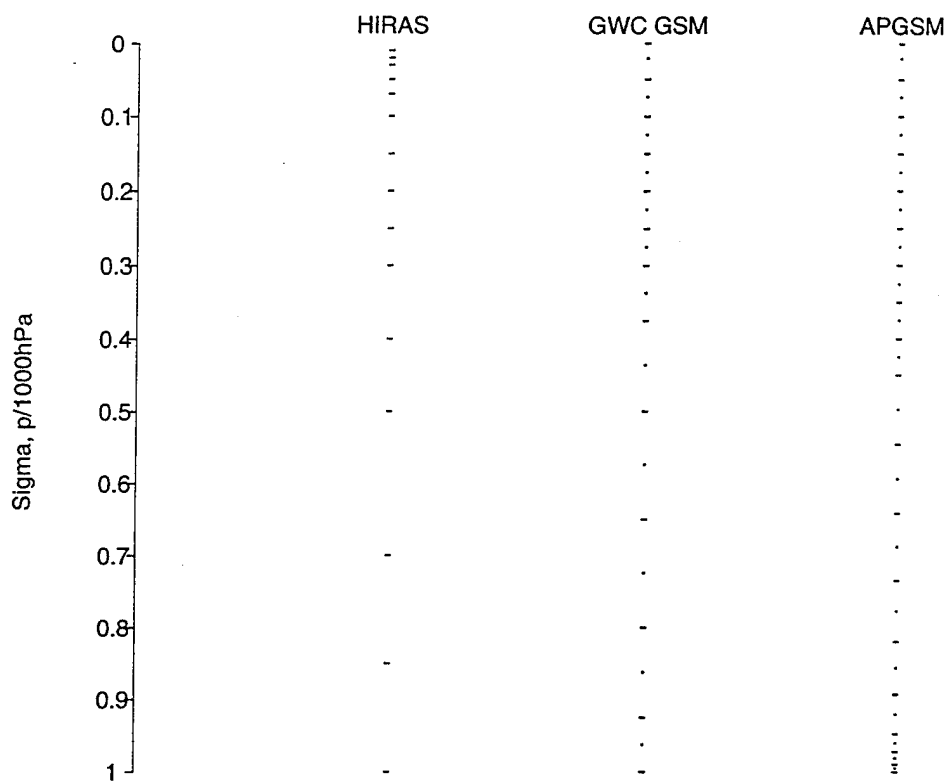


Figure 1: Vertical distribution of the 15 analysis levels (HIRAS) and  $\sigma$ -layers and interfaces for the GWC GSM (12-layers) and APGSM (18 layers).

The parameters of the GSM used in this study closely correspond to those used previously by Nehrkorn et al. (1993a) and Norquist et al. (1992). Specifically, the horizontal resolution is set at rhomboidal truncation 40, and 18 layers are used in the vertical (moisture is carried at all 18 layers). The sigma layer interfaces for the 18 layers are at 1.0, .990, .973, .948, .893, .820, .735, .642, .546, .450, .400, .350, .300, .250, .200, .150, .100, .050, 0. The vertical distribution is also shown in Figure 1. It can be seen that the additional layers are mainly near the ground (to accomodate the planetary boundary layer parameterization), in addition to a slightly higher resolution in the middle troposphere. Above  $\sigma=.3$ , the vertical structure is identical to that of the GWC GSM. A time step of 15 minutes is used, and the coefficients for the time stepping are 0.04 (for the time filter coefficient) and 0.5 (for the time stepping coefficient, implying a semi-implicit step). The coefficient for the horizontal  $\nabla^4$  diffusion is  $2.55 \times 10^{15} \text{ m}^4 \text{ s}^{-1}$ . The terrain data set is a silhouette terrain derived from the Navy 10' dataset; boundary data sets for the physical parameterizations use monthly mean values compiled from a number of sources.

The NMI uses two iterations of the Machenhauer scheme, initializing only the four gravest modes with periods less than 48 hours. For the forecasts of the first half of January, the preprocessor was used unchanged from the version used in the GWC GSM, and the NMI used diabatic tendencies including all physical processes. Because of problems encountered in the initialization (see section 2.3), the preprocessor was modified for later forecasts (the remainder of January, and April, July, and October), and adjustment processes (dry adiabatic adjustment and large-scale precipitation) were excluded from the diabatic NMI.

Throughout its evolution, the PL GSM has undergone testing and comparison with the GWC GSM. Results from tests of a number of preliminary versions can be found in Yang et al. (1989). The most recent version of the physics package was tested by Norquist et al. (1992). They found that for a series of six January and July forecasts (out of 10 days), the APGSM performed better than the GWC GSM. The present study extends these tests in two ways: comparisons with the GWC GSM are performed over a much larger number of forecasts, and the quality of cloud forecasts from the APGSM is studied and compared with operational GWC cloud forecasts.



### 2.3. Comparison Methodology

To evaluate the comparative performance of the GWC GSM and the APGSM for forecasting standard meteorological variables (geopotential heights, winds, and humidity), 1 month of twice daily GSM forecasts out to four days were produced, using the HIRAS analyses as the initial state, and for verification. To assess how much the results vary with season, three additional months (April, July, and October) of forecasts were produced, but only one forecast every three days. This sampling strategy was used to conserve resources. Since atmospheric analyses (and forecasts generated from them) typically exhibit a large amount of serial correlation, this approach does not result in a significant reduction of the effective sample size. The validity of our approach was tested by repeating this sampling procedure for the month of January (see section 3.2). Forecast error statistics were computed over various regions of the globe (see Table 1).

Table 1: List of geographical regions used for the verification of forecasts

Region	Latitude Limits	Longitude Limits
Global	90°S - 90°N	0°E - 360°E
Northern Hemisphere Extratropics	20°N - 80°N	0°E - 360°E
Tropics	20°S - 20°N	0°E - 360°E
Southern Hemisphere Extratropics	20°S - 80°S	0°E - 360°E
North America	25°N - 60°N	120°W - 70°W
Europe	35°N - 70°N	10°W - 40°E

The statistics are the mean (bias), standard deviation, and root mean square, anomaly correlation, and climatological skill score. The latter two are derived with reference to the climatological values for the forecast variables. Anomaly correlation, which is the correlation between forecast and verifying differences from climatological values, is a frequently used measure of forecast model performance and is included here to allow comparison with other operational

models. It can take on values between +1 (perfect) and -1; based on a study of subjectively evaluated forecasts, forecasts with anomaly correlations below 0.6 are generally considered useless. Its use was criticized by Murphy and Epstein (1989) because it does not take into account the mean errors. They proposed instead a climatological skill score, defined as the fractional reduction of the mean square error (MSE) compared to climatology:

$$\text{Skill score} = 1. - \text{MSE (forecast)} / \text{MSE (climatology)}.$$

Skill score values can range from +1 (for a perfect forecast), to 0 (for a forecast no better than climatology), to -n (for forecasts with MSEs (n+1) times larger than that of climatology). To compute the anomaly correlation and climatological skill score we used the Climate Analysis Center (CAC) Climate Diagnostics Data Base (CDDb) 10-year (1979-1988) climatology derived from NMC analyses. This dataset was obtained from the National Center for Atmospheric Research (NCAR). It provides 10-year means of monthly mean values of temperature, geopotential height, and winds, on a 2.5°x2.5° grid, at 8 pressure levels (1000, 850, 700, 500, 300, 250, 200, and 100 hPa).

The initialization procedure for the APGSM underwent two minor changes during the project. The first change consisted of a modification of the preprocessor (relative humidity values above the 300 hPa level are set to a constant value rather than extrapolated) and of the diabatic NMI (adjustment processes are excluded from the tendency calculation). This change was made necessary by unreasonably large initialization increments produced for the 00 UTC 15 January time period over Greenland, which were related to excessive stratiform precipitation rates<sup>1</sup>. It was implemented after the forecasts for January 1-14 had been completed. The second change was necessitated by similarly excessive precipitation rates over Antarctica during April and July. An examination of the problem showed that the temperature structure in the initial state in the top sigma layers over Antarctica led to numerical instabilities involving the radiation and stratiform precipitation parameterization. The 10 hPa HIRAS analyses of geopotential height over that area implied very warm

---

<sup>1</sup>More details can be found in the first Technical Report (Nehrkorn et al., 1993b).

layer mean temperatures, and a steep vertical gradient of temperature. During preprocessing, this led to excessively warm sigma-layer temperatures. We adopted a simple fix to this problem: the 10 hPa level HIRAS geopotential heights over Antarctica were replaced by those computed hydrostatically from the next-lowest (20 mb) layer, assuming an isothermal atmosphere in the top two layers (30mb-20mb and 20mb-10mb). This change was put into place after the January and most of the April APGSM forecasts had been completed.

### **3. Results of Forecast Model Comparison**

#### **3.1. Results for January**

Preliminary results of the forecast model comparison based on partially completed forecasts for the month of January (January 1-20) were reported in Nehrkorn et al. (1993b). Forecast errors averaged over all forecasts in January<sup>2</sup> shown here largely confirm the conclusions from the earlier report. Mean errors of geopotential height at 1000 hPa, 500 hPa, and 300 hPa are shown in Figures 2 - 4 as a function of forecast lead time. In Figure 5, vertical profiles of the bias are shown for all forecast lead times (although the individual curves are not labeled, the different forecast lead times can be identified because the errors generally increase with forecast lead time). It can be seen that in the extratropics, the GWC GSM has an almost linearly increasing positive bias of the 500-1000 hPa and 300-500hPa thickness, whereas the APGSM bias levels off after about 24 hours. This difference is due to the absence of radiative cooling in the GWC GSM (Louis et al., 1989). In the tropics, the APGSM shows a systematic drift towards warmer upper tropospheric temperatures, more so than the GWC GSM. At those latitudes, the moist convection parameterization, and its interaction with parameterized cloud cover and the radiative parameterization, play a dominant role. RMS errors of geopotential height (see Figures 6, 7, 8, for error plots at the 1000, 500, and 300 hPa levels, and Figure 9 for a vertical profile of rms errors) of the APGSM tend to be higher than GWC GSM errors at the beginning of the forecast, but in the extratropics the error growth is smaller in the APGSM forecasts so that the situation there is reversed after 24-36 hours. The improvement of the day 2- day 4 APGSM height errors over those of the GWC

---

<sup>2</sup>Because of logistical difficulties, forecasts were generated for January 1 - 27 instead of 1 - 31.

GSM is particularly pronounced between 500 hPa and 300 hPa (up to 100 hPa in the NH), and also at 1000 hPa. Corresponding plots of the error standard deviation (not shown) exhibit much less consistent APGSM improvements, indicating that most of the improvement is due to the diminished bias. In the tropics, where the APGSM bias is worse than that of the GWC GSM, rms errors are generally worse than those of the GWC GSM, as well. Plots of the anomaly correlation and the climatological skill score for geopotential height (Figures 10 to 13) show that in the extratropics no model is consistently better than the other in terms of the anomaly correlation (as was the case for error standard deviations, which also do not reflect systematic errors), but in terms of the skill score the APGSM is favored at later forecast times (analogous to what was found in terms of the closely related rmse score). In the tropics, height forecasts of either model must be considered without skill after the first 12-24 hours by either the anomaly correlation ( $<.6$ ) or the skill score ( $<0$ ) criterion. Using the same criteria, forecasts in the extratropics remain generally skillful to at least day 3, in most cases through the end of the 4-day forecast.

The horizontal wind vector rms errors (shown in Figures 14 and 15 for the 850 and 300 hPa levels, and as vertical profiles in Figure 16) show the APGSM with consistently larger errors. This degradation compared to the GWC GSM is already apparent in the initial state and it grows during the first 12 hours of the forecast. The degradation is most pronounced at lower levels in the NH extratropics, and in the tropics. The bias of the u and v wind components (not shown) are generally small for either model. The rms errors of u and v (shown as vertical profiles in Figures 17 and 18) are of comparable magnitude, indicating each component contributes roughly equally to the vector errors, and they both show the same qualitative features as the vector rms errors.

The performance of the models in terms of predicting RH is of particular interest in this study because of its obvious relation to cloud forecasts. Both models show some moist spinup behavior. Plots of the bias of RH (Figures 19, 20, 21, and 22) show a tendency of both models to become too moist in the extratropics at 850 hPa and aloft. Most of this moistening takes place during the first 12 - 24 hours of the forecast. It is most pronounced at the upper levels. In general, the RH biases of the APGSM are larger than those of the GWC GSM. Over the North American and European regions, the APGSM bias grows to 20%-

30% by the end of the forecast. In the tropics, both models show a tendency to become too dry (at all levels), the APGSM more so (up to 10%) than the GWC GSM. In terms of RH rms errors (Figures 23, 24, 25, and 26), the GWC GSM forecasts are better at the beginning of the forecast at all levels, and for all regions. After the first 12-36 hours of the forecast, this continues to hold true in the tropics, and at the upper levels (500 hPa and above) in the extratropics. At 1000 and 850 hPa in the extratropics, APGSM rms errors grow more slowly and are smaller than those of the GWC GSM in the latter half of the forecast.

Based on the forecast results for the month of January, the APGSM shows only one area in which it consistently improves on the performance of the GWC GSM, namely the warm bias of the extratropical troposphere (in the tropics, it has a larger warm bias than the GWC GSM in the upper troposphere). In practically every other measure of skill, APGSM results are worse than those of the GWC GSM during the first 1-2 days of the forecast. At later forecast lead times, results are mixed. In the extratropics, the APGSM has smaller RMS errors of geopotential at the upper levels, and of RH at 1000 and 850 hPa. RMS errors of the horizontal wind (vector errors and u and v components) are generally larger for the APGSM, as are the RH bias and RMS errors at upper levels. In general, the degradation of the APGSM performance is worse in the tropics. The relatively poor performance of the APGSM may be partly due to the verification procedure employed here, and partly due to model deficiencies. The use of HIRAS analyses put the APGSM at a potential disadvantage because (1) no special effort was made, aside from straightforward preprocessing and initialization procedures, to minimize errors in the initial state and resulting spinup problems; and (2) because the GWC GSM is used as the first guess model for the HIRAS analysis, and might thus bias the analysis in data sparse regions toward the GWC GSM. The fact that APGSM errors are larger than GWC GSM errors initially, but then become smaller in certain areas, suggests that efforts at improving the initialization procedure might prove beneficial. The argument that HIRAS analyses favor the GWC GSM might apply to some degree in the tropics and SH, but at least in the NH extratropics, comparison of the well-sampled North American and European regions with the entire NH extratropical latitude belt shows no systematic bias of the results toward either model. It thus appears that model deficiencies are at least partially to blame for the larger APGSM errors. Since wind errors, in particular at low levels in the NH

extratropics, are disproportionately worsened, factors that selectively affect the winds, such as the friction parameterized in the PBL, or the use of silhouette orography and gravity wave drag, might warrant further study. The forecast errors in the tropics indicate, to the extent that the HIRAS analyses can be trusted there, that the convection scheme, and/or its interaction with the radiation scheme, need to be tuned.

### **3.2. Results for Subsets of January Forecasts**

To assess the effect of sampling strategies on the results of the forecast model comparisons, we recomputed the January verification statistics for a subset of forecasts that replicates the approach taken for the other months: forecasts from every third day, starting at 00 UTC January 1, were used in one such comparison. Another subset was selected by starting from 12 UTC 2 January. The results showed no appreciable differences between the different samples, aside from a slightly noisier appearance of the error curves. A similar noisiness is also apparent in the curves for April, July, and October (see next section). It is a consequence of the smaller sample sizes.

### **3.3. Results for April, July, and October**

The variation of the forecast model performance with the seasonal cycle was assessed by a series of 4-day forecasts, spaced 3 days apart, in the months on April, July, and October. Vertical profiles of geopotential height bias, shown in Figures 27 - 29 for those three months, show generally similar results to those of January (viz. Figure 5) in the SH extratropics and the tropics. In the NH extratropics, the GWC GSM bias is somewhat smaller in April and July, and the APGSM bias is larger in July above 300 hPa (most likely because this region is more affected by convective processes during this season). The RMS errors of geopotential height (Figures 30 - 38) show only minor differences from the results in January. In particular, the better APGSM scores at 1000 hPa seen in January in the extratropics are absent in the NH extratropics in October, and in all three months SH APGSM errors are worse than those of the GWC GSM at 1000 hPa. At the upper levels, the error saturation of the GWC GSM occurs at smaller (larger) values in the NH (SH) extratropics in July, resulting in a smaller improvement of the APGSM relative to the GWC GSM. Thus, APGSM rms height errors are improved the most in the winter hemisphere, at the jet level, toward the end of the 4-day forecast. In the tropics, APGSM errors tend to be

larger in April and July. RMS errors of the horizontal wind (Figures 39 - 41) show generally the same picture as those for January (viz. Figure 16), with some minor exceptions. In some circumstances, APGSM wind errors are slightly smaller than those of the GWC GSM (in July, at day 4, in the SH extratropics above 300 hPa; in April and October, at day 3-4, in Europe above 300 hPa). The bias and RMS of RH (not shown) are generally quite similar to the January results.

#### **4. Cloud Forecasts and Observations**

A number of different schemes to diagnose clouds from the model output were investigated. Aside from the two schemes that were included for reference purposes only (persistence, and the Slingo scheme used in the radiation parameterization of the APGSM), there are two basic categories: regime-based schemes, and regression schemes. Both categories attempt to relate cloud cover not only to RH, as is primarily done in the Slingo scheme and numerous other comparable operational schemes, but also to other model variables. The regime based schemes use a classification of vertical profiles of atmospheric variables into synoptic regimes to stratify cloud cover statistics, whereas the regression-based schemes use multiple linear regression between cloud amount and various atmospheric parameters. Both schemes require observations of cloudiness for the derivation of the proper statistics. In the following, we describe the observational cloud data base used in the derivation of both types of schemes, and then go on to describe the different cloud forecast scheme.

##### **4.1. RTNEPH Data**

Cloud data were needed for the development and testing of the cloud forecast schemes. For this purpose, Northern Hemisphere RTNEPH (Hamill et al., 1992) data (for the four months for which forecasts were produced) were used. The raw data consisted of total cloudiness, layer cloud cover of up to four layers, and various data source, timeliness, and cloud type information on the so-called 1/8-mesh grid (this is a regular grid on a polar stereographic projection, with a grid spacing of 47.625 km at 60° N). Before these data were used in our study, they were compacted to the 1/2-mesh (grid spacing of 190.5 km at 60° N), so that they represent scales resolved by the forecast model. Data at this resolution are also used by GWC in the verification of the operational cloud forecast models.

The 1/8-mesh RTNEPH data were unpacked and compacted to the 1/2-mesh all in one step. The final output from this compaction were working sums for averaged cloud cover (total and 6 layers). The working sums consist of the sum of weights and the sum of weight times cloud cover.

The methodology was as follows: The up to 4 floating RTNEPH layers were assigned to 6 fixed MSL layers. The MSL layer tops were chosen to correspond approximately to the 6 mandatory pressure levels between 1000 hPa and 300 hPa, except that the top layer includes all high clouds. Layer boundaries are at 0, 1.07, 1.98, 4.27, 6.71, 7.92, and 25 km. Working sums for horizontal averages for the total cloud amount, and the 6 MSL layer amounts, were then formed by using a 25-point weighted average with a 1-2-2-2-1 weighting applied in both the i and j directions. (The weights are between 1 and 4, and the maximum for the sum of weights is 64.) Layer clouds identified as thin were set to zero cloud cover. If more than one floating layer contributed to a fixed MSL layer at one 1/8-mesh point, maximum overlap was assumed. Only points with valid cloud cover data satisfying the timeliness criterion (data used for cloud analysis were within +/- 2 hours of the valid time) were used. Based on the fact that most half-mesh points either have no data (sum of weights=0) or complete data (sum of weights=64)<sup>3</sup>, the compacted data were used only at those gridpoints where averages could be formed with complete 1/8-mesh data.

Use of the RTNEPH data in conjunction with the forecast model requires coordinate transformations between the 1/2-mesh grid, which is a regularly spaced grid on a polar stereographic projection, and the grids used by the forecast model and analysis, which are regularly spaced in longitude, at either regularly spaced (analysis) or Gaussian (GSM transform grid) latitudes. Routines existed to convert longitude/latitude coordinates to coordinates in the polar stereographic projection; however, for the purpose of converting grid-box average quantities, the locations of all lat-lon grid boxes overlapping a given 1/2-mesh box were needed. Lookup tables were generated that contain, for each RTNEPH point, the locations of all GSM transform grid (128 longitudes by 102 Gaussian latitudes) boxes overlapping the RTNEPH grid box, and the fraction of

---

<sup>3</sup>More details can be found in the first Technical Report (Nehrkorn et al., 1993b).



area covered by the RTNEPH grid box. RTNEPH cloud cover can thus be properly transformed to the GSM transform grid by forming weighted averages, and the reverse process can be used to project forecast cloud cover from the GSM transform grid to the RTNEPH grid.

## **4.2. Regime-based Cloud Forecast Schemes**

We developed a number of different forecast schemes based on statistics that were stratified by synoptic weather regimes. The simplest one of these used just the average cloud cover of each regime as its forecast (Zivkovic and Louis, 1992). All others make use of the Cloud Curve Algorithm (CCA) scheme (Mitchell and Hahn, 1989), which consists of a mapping of forecast relative humidity to observed cloud cover, derived from the cumulative probability distribution of each variable. The definition of the weather regimes, and the derivation of the cloud statistics for the different regimes, is discussed in the following.

### **4.2.1. Definition of weather regimes**

The synoptic weather regimes were derived from a statistical analysis of the HIRAS analysis data for the Northern Hemisphere. For this purpose, the HIRAS analyses were preprocessed, i.e. vertically interpolated to the GSM  $\sigma$ -layers, and spectrally truncated to R40; this was done to allow easier application of the synoptic weather regimes (which are defined in terms of vertical profiles of atmospheric quantities) to the GSM output. The preprocessed analyses were sampled at locations spaced approximately 1000 km apart in the Northern Hemisphere, for four time periods, spaced one week apart, in each of the four months (January, April, July, and October). To reduce the number of degrees of freedom, the vertical profiles of all atmospheric quantities (temperature, wind speed, relative humidity), along with the values of surface pressure and precipitable water, were represented by empirical orthogonal functions (EOFs), and weather regimes were then defined in terms of clusters of EOF coefficient values. A detailed discussion of the EOF and cluster analysis is presented in Appendix A.

Of the different clustering procedures discussed in Appendix A, the "average-out" method was used in an initial clustering of the entire sample of

4672 elements. The procedure was based on loading scores of the first 8 principal components. It resulted in 36 clusters, of which only 16 were of significant size to be retained for further analysis. The total number of neglected observations was 97 (approximately 2% of the total sample). The second row in Table 2 lists the number of elements within each of the retained 16 clusters. Clusters are ordered and numbered according to a decreasing value of the lowest sigma-layer temperature, that is the warmest cluster (near the surface) is marked as cluster 1. Cluster 16 is the coldest cluster (near the surface). It is noticeable that among the 16 clusters, one is substantially larger than the others. It contains 2352 observations (elements) that are located mostly in the tropical region. This tropical cluster was subdivided in a second step using the "average-in" method (it was found to provide more diverse mean values in the tropics). The second clustering step was terminated when 13 clusters were formed. Only 9 of the clusters had more than 50 elements and were retained for the final analysis. Their number of elements is listed in the last row of Table 2. A total of 122 observations was neglected this way which accounts for approximately 2.6% of a total sample. Hence, during the two steps of the clustering analysis a total of 219 observations were neglected from the 4672 original observations (4.7%).

Table 2. Number of elements in retained clusters obtained in two step clustering procedure.

	Cluster Number															
Step	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Step 1	2352	15	399	226	14	66	17	407	124	292	132	77	45	15	378	16
Step 2	513	152	209	193	405	481	120	105	52	-	-	-	-	-	-	-

A total of 24 clusters, or weather regimes, was generated in Steps 1 and 2. Clusters and associated regimes are ordered according to a decreasing value of temperature variable near the surface. Table 3 summarizes the resulting numeration of regimes. Clusters are identified by appropriate numbers from Table 2, which are listed in columns 2 and 3. The number of observations within each regime is given in column 4. Clusters are not uniform in size but most of them are large enough to provide stable mean values. These mean profiles of the regimes are plotted in Figure 42. Note that these profiles are computed directly from the atmospheric profiles of the observations, and not from a truncated projection onto the EOFs.

Table 3. Numbering of weather regimes after clustering steps 1 and 2.

Regime number	Step Cluster Number		# Observations
	Step 1	Step 2	
1	-	1	513
2	-	2	152
3	-	3	209
4	-	4	193
5	-	5	405
6	-	6	481
7	2	-	15
8	-	7	120
9	-	8	105
10	3	-	399
11	4	-	226
12	-	9	52
13	5	-	14
14	6	-	66
15	7	-	17
16	8	-	407
17	9	-	124
18	10	-	292
19	11	-	132
20	12	-	77
21	13	-	45
22	14	-	15
23	15	-	378
24	16	-	16

When applying the synoptic climatology derived above to arbitrary analysis or forecast fields, a given grid point is assigned to the regime with the most similar profile of the 38 state variables (temperature, wind speed, relative humidity, and surface pressure and precipitable water). Similarity is simply measured by the Euclidean distance of the normalized profiles. The definition of the weather regimes was modified after it was tested on forecasted fields, because the skill of the model in predicting the correct regime was quite poor for some of the 24 regimes identified above. A contingency table of forecasted vs. verifying weather regime was examined for a one week period in April (00 UTC 5 April to 12 UTC 11 April). The standard contingency table scores Critical Success Index (CSI), Probability of Detection (POD), and False Alarm Rate (FAR) were computed for each of the 24 regimes, and for a number of possible combinations of regimes. The scores for a given regime (r) can be defined as follows:

$$\text{CSI} = A / (A+B+C)$$

$$\text{FAR} = C / (C+A)$$

$$\text{POD} = A / (B+A)$$

where A represents the number of gridpoints where both the regime diagnosed from the forecast (f) and the verifying analysis (v) are equal to r ("hits"); C represents "false alarms" ( $f=r, v \neq r$ ); B represents "misses" ( $f \neq r, v=r$ ). The forecast and verifying sample sizes are then given by (A+C) and (A+B), respectively. Scores for CSI, FAR, and POD range from 0.0 to 1.0; a high CSI and POD is desirable, as is a low FAR. The CSI gives the best overall measure of skill, since it takes both the hits and misses into account.

Deciding which and how many regimes to combine involves a compromise between homogeneity of the regime (and associated cloud and RH statistics) on the one hand, and skill in predicting the regime on the other hand. This was

done subjectively, using the mean regime profiles<sup>4</sup>, sample sizes, and contingency table scores of the individual and combined regimes as factors in the decision making. The resulting final list of regimes is shown in Table 4. It consists of 7 combined regimes (with up to 4 individual regimes in one combination), and 8 individual (original) regimes, for a total of 15 regimes (compared to the original 24 of Table 3). Figure 43 shows the sample sizes and contingency table scores of the combinations and individual regimes, for forecast lead times from 12 hours to 96 hours. A comparison of the verifying and forecast sample sizes (note that verifying sample sizes appear as flat lines in Figure 43, since they do not depend on forecast lead time) gives an indication of the systematic underforecast (e.g., regime 5) or overforecast (e.g., regime 11). For some of the combined regimes, opposing biases of the member regimes lead to a decreased overall bias (combinations 11\* and 23\*). As can be seen from the CSI scores, regimes were only combined if the combined CSI was better than the CSI of any of its members. The dependence of skill on forecast lead time is quite different for the different regimes. For the tropical regime 1\* there is very little dependence on forecast lead time (the skill is uniformly low), most likely because error levels in the tropics are already near their saturation level early on in the forecast (viz. section 3). For the high latitude regime 23\* the skill remains quite high, because the fractional change in area covered by the regime is small over the length of the forecast (and, in addition, forecast skill at high latitudes is better than in the tropics). Regime 19\*, on the other hand, shows a strong drop-off in CSI with forecast lead time, reflecting the decrease in skill of the placement of the midlatitude jet stream with which this regime is associated.

We studied the geographical distribution of the regimes for a number of preprocessed HIRAS analyses (00 UTC 1 January, 12 UTC 3 January, 00 UTC 5 January, 00 UTC 1 April, 12 UTC 3 April, and 00 UTC 5 April). Overall, we found a clear correspondence between the weather regimes and synoptic and/or geographic features. The most important factor controlling the geographical distribution of the regimes is the position of the major baroclinic zone in the

---

<sup>4</sup>Combination of the regimes does not involve computation of new mean profiles for the regimes; instead, the gridpoints are first assigned to one of the 24 original regimes, and then reassigned based on the list of combined regimes.

midlatitudes. This is clearly evident in Figure 44, which shows an example of the geographic distribution of the regimes together with contours of the 1000hPa - 500 hPa thickness field. The cold-air regimes (21\*, 23\*) can be found to the north of this zone, midlatitude regimes (7, 13-19\*) are generally along the strongest temperature gradients, and subtropical (8\*, 10, 11\*) and tropical (1\*, 2\*) regimes are to the south. Two regimes are clearly associated with geographic features: regime 21\* is a mountain regime of middle and high latitudes (including the Himalayas, parts of the Rockies, and Greenland); regime 10 is a dry regime found over low-latitude land areas (including the Sahara, parts of East Africa, and the Saudi-Arabian peninsula).

The breakdown of these major groupings of regimes can be related to a number of different meteorological features. In high latitudes, the three regimes differ mainly by the relative humidity. Midlatitude regimes are broken down by their relation to wind speed minima and maxima, both at lower and upper levels, and by differences in the humidity structure. For example, Figure 45 shows that regimes 17, 18, and 19\* are all associated with the midlatitude jet stream. The regimes can sometimes also be related to synoptic features in the geopotential fields. As one goes further south, the relation with synoptic features in the height and wind speed becomes less well defined, until in the tropical clusters the only discernible features related to the regime distribution can be found in the humidity fields. A brief summary of synoptic and geographic features for each of the 15 regimes is included in Table 4. The relationship of these regimes with cloud cover will be explored in greater detail in the following two subsections, in which the cloud cover and RH statistics are investigated separately for the different regimes.

Table 4: List of the final 15 weather regimes. First column is the regime number ("\*"s indicate combined regimes). Second column lists the regime numbers from the original 24 regimes included in the combined regimes.

No.	Comb.	Description
1*	1,6	Moist tropical regime
2*	2,3,4,5	Dry tropical regime
7		Oceanic midlatitude regime, at southern edge of subtropical 300 hPa jet
8*	8,9	Subtropical regime with weak flow at all levels
10		Dry, low-latitude regime found over land (including deserts)
11*	11,12	Subtropical regime with weak flow at all levels, pronounced vertical gradient of RH. Associated with surface highs.
13		Midlatitude regime found near subtropical 300 hPa jet core, in areas of weak surface flow.
14		Midlatitude regime found near subtropical 300 hPa jet core, in areas of weak surface flow.
15		Midlatitude regime with high RH at all levels. Found in areas with strong low-level warm and moist advection, removed from the 300 hPa jet core.
16		Midlatitude regime found in areas with weak surface flow, removed from 300 hPa jet cores.
17		Midlatitude jet stream regime.
18		Midlatitude jet stream regime
19*	19,20	Midlatitude jet stream regime
21*	21,22	Dry mid-high latitude regime associated with high elevations
23*	23,24	High-latitude regime, found north of the main baroclinic zone

#### 4.2.2. Average cloud cover schemes

The simplest possible parameterization of cloud cover based on the weather regimes is to assign each regime its average cloud cover. This was done in Zivkovic and Louis (1992), and it has been included here to separate the contributions from the regime classification and the CCA curves discussed in the next section. The average cloud cover at the 6 RTNEPH layers, and for total cloud, are given in Table 5. These values were derived from one week (00 UTC 5 April to 12 UTC 11 April) of RTNEPH data and preprocessed HIRAS analyses. The April time period was chosen for an even distribution of weather regimes

(summer or winter time periods would have disproportionately smaller or larger number of cold weather regimes). The two columns for total cloud cover correspond to the value obtained from the reported total cloud cover, and from the layer cloud cover (using the overlap procedure described in Appendix B), respectively. Both methods for specifying total cloud cover were tested using forecasts for the one-week data, with only minor differences between them. For consistency with the CCA schemes discussed below, the overlapped total cloud amount was chosen for tests on independent data.

Table 5: Average cloud cover for the regimes, for the six RTNEPH layers, and for total cloud cover from the RTNEPH data ( $T_r$ ) and from the overlap computation ( $T_o$ ).

Regime	Layer						$T_r$	$T_o$
	1	2	3	4	5	6		
1*	28.6	21.2	18.4	9.7	5.8	5.1	38.1	39.16
2*	32.8	32.9	39.7	33.4	22.7	19.1	64.2	59.26
7	34.5	33.8	42.5	34.1	21.0	16.4	70.3	61.16
8*	23.7	19.5	17.2	8.9	5.3	5.0	33.3	33.82
10	4.1	5.7	8.7	5.3	2.6	4.6	14.2	14.14
11*	32.1	25.3	21.6	9.2	4.2	3.6	44.3	42.03
13	20.5	25.2	41.6	35.7	19.5	25.8	64.6	59.74
14	14.7	15.4	23.0	16.7	8.3	6.9	35.0	34.68
15	30.6	35.2	57.7	62.9	39.8	28.0	89.4	80.47
16	32.8	36.7	45.3	22.5	8.7	8.1	61.9	58.19
17	40.5	40.3	46.9	27.9	13.6	10.1	70.4	61.39
18	27.1	29.7	36.6	18.6	7.4	7.0	52.5	49.02
19*	24.7	31.8	52.3	39.1	16.8	11.8	70.0	67.04
21*	3.8	8.1	33.3	27.1	7.7	8.5	42.0	43.05
23*	5.3	10.2	31.1	15.5	3.1	2.9	36.0	38.19

#### 4.2.3 Cloud curve algorithm schemes

The CCA schemes developed in this project refine the simple average cloud cover scheme discussed in the previous section by explicitly modeling the relation between RH and cloud cover within each regime. A mapping between



RH and cloud cover is derived by matching RH and cloud cover values corresponding to equal cumulative probability functions (Mitchell and Hahn, 1989). Thus, for each regime and each cloud layer, histograms of observed cloud cover and forecast RH have to be compiled. There are several technical details that must be addressed in the implementation of the technique:

(1) Effects of regime forecast errors:

Since the 15 synoptic regimes correspond to synoptic scale features, forecast errors will, in general, result in errors in the geographical placement of the regimes (see section 4.2.1). These errors introduce difficulties when forecast RH has to be related to observed cloud cover for each regime. If the forecast is used to define the regime for a given location and time, then the cloud cover statistics for the regimes will not be those of the true regimes; if the analysis is used for regime definition, the RH statistics will not be those of the forecast regimes. We chose instead to use the forecast to define the regime for the forecast RH, and the analysis for the observed cloud cover. In this approach, histograms of observed cloud cover and forecast RH are those of the respective regimes; however, the statistics will not only be those of colocated points.

These difficulties are not unique to our derivation of the CCA curves. If relations between forecast RH and cloud cover are derived from colocated RH forecasts and observed cloud cover, forecast errors in RH introduce scatter and contaminate any relation between the true RH and cloud cover. If CCA curves are stratified by geography rather than regime, the CCA curves are contaminated to the extent that placement errors of RH features are large enough to affect the RH statistics for a given region.

(2) Remapping of cloud and forecast data

Because cloud and forecast data are available at different horizontal and vertical grids, they need to be remapped to a common coordinate system. In the horizontal, the RTNEPH data were remapped from the RTNEPH grid to the GSM transform grid. In the vertical, the GSM sigma layer RH data were remapped to the six constant-pressure layers of the RTNEPH by choosing either the maximum or average RH of the appropriate sigma layers. Separate CCA

curves were derived for average and maximum RH values, and both variants were tested on the dependent data set.

### (3) Prediction of total cloud cover

The total cloud cover was computed from the predicted layer cloud amounts, using the procedure described in Appendix B, for one version of the CCA schemes. Because total cloud cover is generally regarded the most reliable RTNEPH analysis variable, an alternative scheme was implemented in which separate CCA curves were derived for total cloud cover. In this approach, total cloud cover statistics were related to column-average and column-maximum (between  $\sigma=1$  and  $\sigma=.3$ ) RH, respectively. To ensure consistency between total and layer cloud amounts, the *layer* cloud amounts were adjusted such that the total cloud cover was no larger than what would result from random overlap, and no smaller than what would result from maximum overlap.

Examples of the cloud cover histograms obtained from the one-week period in April are given in Figures 46, which shows total cloud cover statistics. The histograms show substantial variation from one regime to another, with maxima at clear and/or overcast values. The mean values correspond to those of Table 5. Corresponding histograms were derived for the layer cloud cover, and the forecast RH. From these histograms, cumulative probability density functions and CCA curves were computed for four separate cloud forecast schemes, using two methods of remapping sigma-layer RH (average and maximum RH) and of deriving total cloud cover (overlapping and separate CCA curve). Figure 47 shows the resulting CCA curves for 12 hour forecasts, using the average RH of the cloud layers. The corresponding curves for 24 hour forecasts are shown in Figure 48. Differences between the curves at different forecast times reflect the effects of model spinup and drift: changing systematic errors in RH will result in systematic shifts of the CCA curves towards lower or higher RH values, minimizing the bias in forecast cloud cover. For example, the CCA curves for regimes 13-20\*, particularly at the upper levels, are shifted toward higher values of RH, compensating for the trend of the model to moisten (viz. section 3). At later forecast times, differences between CCA curves are more gradual, which is why CCA curves were combined for forecast lead times of 36-48 hours and 60-96 hours (not shown).

All four variations of the CCA scheme were tested on the dependent data set, alongside the two average cloud cover schemes, in order to identify the most promising candidate for extensive testing on independent data. For verification purposes, cloud forecasts were compared to the RTNEPH data on the RTNEPH grid, over the octagon also used operationally by GWC (see Figure 49). Separate verification statistics were also accumulated for the different regimes, on the GSM transform grid. The results for the octagon are shown in Figures 50, 51, 52, and 53 for 12 hour forecasts. The results from the Slingo cloud scheme (discussed below) are included as a reference. Ideally, since the CCA curves are designed to result in identical probability density functions of forecast and observed cloud amounts, both bias and sharpness measures of the CCA schemes should be nearly perfect, at least on the dependent data set. Departures from perfect scores seen in Figures 50 and 51 are due to the fact that the placement of the regimes in the forecasts is in error. The systematic errors are generally small (less than 5%) for the CCA schemes for all layers, as is to be expected. Total cloud cover biases are somewhat larger for the schemes using overlap. The sharpness, defined as the percentage of forecasted cloud amounts within 20% of clear or overcast, tends to be slightly less than that of the RTNEPH data for the CCA schemes. Using separate CCA curves for total cloud cover, and adjusting layer cloud covers as needed for consistency, tends to sharpen the forecasts of both total and layer cloud amounts. The Slingo scheme tends to produce forecasts that are too sharp, whereas the average cloud cover schemes, not surprisingly, are much too smooth (except at the upper levels, where average cloud covers are mostly below 20% - see Table 5). The root mean square error and correlation scores shown in Figures 52 and 53 show that, in general, the forecasts with the lowest sharpness also tend to produce the smallest errors in a mean square sense. Somewhat surprisingly, the average cloud cover schemes are best overall in terms of the rmse and correlation. While one might argue that the low rms errors of the average cloud cover schemes are due to the fact that they are overly conservative, their higher correlation scores indicate that the different regimes resolve all the predictable variance of the cloud cover, and attempts to sharpen the forecasts within the regimes using the CCA schemes are, on average, detrimental. One might question whether some of the small differences between the schemes are truly representative. An indication that this is so is shown in Figure 54, which shows a scatter plot of rms scores of total cloud for the average RH schemes using overlap and a separate CCA curve. Although the mean rms

scores for the 14 12-hour forecasts differ by less than 3%, the overlapping scheme clearly has consistently better scores. The results shown in Figures 50 - 53 for 12 hour forecasts generally hold true at later forecast times, as well. Figures 55 - 57 show the bias, rmse, and correlation scores for 96 hour forecasts.

The results of the cloud scheme tests on the dependent data set do not identify a clear winner, since the schemes with the lowest rms errors (the average cloud cover schemes) have highly unrealistic cloudiness distributions. We decided to select the CCA scheme with the lowest rms error (using average RH, with total cloud computed from overlapping layer amounts), since it represents a compromise between the schemes with the best sharpness and best rms error. As a reference, the average cloud cover scheme was also retained; for consistency, this scheme also uses overlapping for the prediction of total cloudiness.

#### **4.3. Regression Schemes**

A multiple linear regression of cloud amount vs. a large number of atmospheric parameters computed from forecast model output was developed by PL (Norquist et al., 1994). Cloud amount in three terrain-following layers (denoted high, middle, and low cloud decks) was regressed against a total of 98 possible predictors. The cloud decks covered the vertical intervals from  $\sigma=1.0$  to  $\sigma=0.8$  (low),  $\sigma=0.8$  to  $\sigma=0.45$  (middle), and  $\sigma=0.45$  to  $\sigma=0.20$  (high). The predictors for each deck consisted of some geographic and solar parameters (e.g., terrain, land/water fraction, hours of sunshine or darkness preceding the forecast), column-integrated or surface variables predicted by the forecast model (e.g., surface pressure, precipitation), and a large number of parameters that were maximum or average quantities for the sigma-layers spanning the cloud decks. Both forecasts verifying at the valid time, and 6 hours earlier, were used as predictors. Some deck-average predictors were used only at the predictand deck (e.g., vorticity advection), whereas for others (e.g., vorticity) the averages for all three decks were considered as predictors for each of the three cloud deck amounts. Table 6 provides a summary of the predictors from the forecast verifying 6 hours before the valid time, and 7 the predictors at the valid time. The IMSL library routine for forward stepwise regression (RSTEP) was used, with a maximum of 20 predictors allowed in each regression equation. Standard values were used for stopping criteria (the p-value for entering variables into the equation was set at 0.05, and the tolerance for detecting linear dependence

among predictors was set at 100 times the relative machine precision), but the sample size parameter was set to  $1/10^{\text{th}}$  the actual sample size to account for the fact that the samples were not truly independent. In Norquist et al. (1994), the APGSM was run with a horizontal resolution of triangular 106 (T106) and 22 layers, using ECMWF analyses for 1991 for the initial state, and  $1/8^{\text{th}}$  mesh RTNEPH data remapped to the GSM horizontal and vertical grid. Forecasts of the previous 10 days were used to compute the regression equations.

Table 6: PL scheme predictors from the forecast valid 6 hours before the cloud forecast valid time. The column labeled No contains the designation used by Norquist et al. (1994). Unless otherwise noted, all quantities are averages of the predictand deck. Adapted from Norquist et al. (1994).

No	Description
1	Vorticity
2	Divergence
3	Temperature
4	Precipitable water
5	RH
6	Vertical velocity
7	$d\theta/dz$ (vertical gradient of potential temperature)
8	Wind speed
9	Wind shear
10	Vorticity advection
11	Temperature advection
12	Three-dimensional humidity advection
13	Condensation pressure deficit
14	$d\theta_e/dz$ (vertical gradient of equivalent potential temperature)
15	$u$ (zonal wind)
16	$v$ (meridional wind)
17	Maximum RH within predictand deck
18	RH at layer above the layer with the maximum RH
19	Temperature at the layer with the maximum RH
20	$d\theta/dz$ at layer above the layer with the maximum RH
21	Surface pressure
22	Stratiform precipitation (at the surface) over the last 6 hours
23	Convective precipitation (at the surface) over the last 6 hours
24	Evaporation (at the surface) over the last 6 hours
25	Surface layer wind speed
98	$RH^2$
99	$RH^4$

Table 7: Predictors valid at the verifying time used in the PL scheme. Unless otherwise noted, deck averages are used. Column labeled Deck indicates whether quantities are for all 3 decks (3), the predictand deck only (1), surface (S) or column-integrated (C). A "Y" entry in the column labeled Res indicates predictors used in the regression of the residuals from the average cloud cover scheme. Adapted from Norquist et al. (1994).

No	Deck	Res	Description
26-28	3	Y	Vorticity
29-31	3	Y	Divergence
32-34	3	Y	RH
35-37	3	Y	Vertical velocity
38-40	3	Y	$d\theta/dz$ (vertical gradient of potential temperature)
41-43	3		wind speed
44-46	3		wind shear
47-49	3	Y	Maximum RH
50	1		Temperature
51	1	Y	Precipitable water
52	1	Y	Vorticity advection
53	1	Y	Temperature advection
54	1	Y	Three-dimensional humidity divergence
55	1	Y	Condensation pressure deficit
56	1	Y	$d\theta_e/dz$ (vertical gradient of equivalent potential temperature)
57	1		u (zonal wind)
58	1		v (meridional wind)
59	1		RH at layer above layer with maximum RH
60	1		Temperature at layer with maximum RH
61	1		$d\theta_e/dz$ at layer with maximum RH
62	S		Surface pressure
63	S	Y	Stratiform precipitation over last 6 hours
64	S	Y	Convective precipitation over last 6 hours
65	S	Y	Evaporation over last 6 hours
66	S		Surface layer wind speed
67	1		Maximum RH <sup>2</sup>
68	1		Maximum RH <sup>4</sup>
70	1		RH with respect to ice at layer of maximum RH

Table 7 (continued)

No	Deck	Res	Description
71	1		Lifted condensation distance at layer of maximum RH
72	1		Ln(Ri) (Log of Richardson No) at layer of maximum RH
73	S		Latitude
74	S		Sine of latitude
75	S		Cosine of latitude
76	S		Sine of longitude
77	S		Cosine of longitude
78	S		Solar zenith angle at initial time of forecast
79	S		Cosine of solar zenith angle at initial time of forecast
80	S		Hours of sunshine before initial time of forecast
81	S		Hours of darkness before initial time of forecast
82	S		Surface terrain height (average of closest 9 1/8 <sup>th</sup> -mesh points)
83	S		Surface terrain height (std dev of closest 9 1/8 <sup>th</sup> -mesh points)
84	S		Percent of surface covered by water
85	S		Eastward gradient of terrain height
86	S		Northward gradient of terrain height
87	C		Minimum of ln(Ri) over 9 transform grid points
88	C		Minimum of $d\theta/dz$ over 9 transform grid points
89	C		Maximum of wind shear over 9 transform grid points
90	C		Maximum of wind speed over 9 transform grid points
91	S		Maximum of convective precipitation over 9 transform grid points
92	S		Maximum of surface layer wind speed over 9 transform grid points
93	S		Maximum of (surface wind x terrain variance) over 9 transform grid points
94	1		Wind times terrain gradient
95	C		Smaller of terrain height or wind/stability height
96	1	Y	RH <sup>2</sup>
97	1	Y	RH <sup>4</sup>



This scheme was implemented in the current study with some minor modifications to accommodate the different grid structures. In particular, the compacted RTNEPH data (on the 1/2-mesh grid and on the six layers) was remapped to the APGSM R40 transform grid, and, using forecast surface pressure, remapped to the three terrain-following cloud decks. Maximum overlap of the pressure-layer cloud amounts was assumed for this remapping. Total cloud cover was computed from the three deck amounts using the overlapping procedure described in Appendix B. The covariance statistics of the predictors and RTNEPH cloud cover from the forecasts of the previous 10 days were used as input to the regression scheme for each forecast, analogous to the procedure used by Norquist et al. (1994).

In addition to the PL scheme, we implemented a regression scheme designed to resolve the within-regime variance of cloud cover. In this scheme, the residuals from the average cloud cover scheme described in section 4.2.2 were regressed against a subset of the predictors used in the PL scheme. Because of logistical constraints, global regression equations were developed rather than separate ones for each regime. The cloud amount predictands were those on the 6 RTNEPH layers, and the predictors were the APGSM forecast variables of the three terrain-following decks so marked in Table 7. Predictors such as wind speed and wind shear, and geographic and solar parameters, were excluded from the pool of predictors because they had already been used in the definition of the regimes, and so presumably were already accounted for.

The regression equations for the PL scheme almost always contained the maximum number of predictors (from 15 to 20, 19.98 on average), and multiple  $r^2$  values ("fraction of explained variance") ranged from 12.6% to 35.2%, 26.5% on average. A list of the most frequently selected predictors (those selected in at least 1/3 of cases) is given in Table 8. Not surprisingly, measures of relative (variables 32-34) and absolute moisture (variable 4) are among this list, as are measures of static stability (variables 7, 40, 88). A large number (8) of location-dependent parameters (variables 21, 74-77, 82-84) are among this list, accounting for the time-mean cloudiness distribution. With very few exceptions (variables 1, 85, 86, 94), all variables were used at least at some point, indicating frequently changing regression equations. This resulted from the use of a large pool of predictors, many of which are correlated with each other. Regression equations

for the average cloud cover scheme residuals contained fewer predictors (4 to 19, 10.08 on average), and multiple  $r^2$  values were in the range 0.9% to 18.8%, 7.0% on average. The lower  $r^2$  values reflect the fact that a large amount of the cloud cover variance is already accounted for by the synoptic weather regimes. The corresponding list of most frequently selected predictors is given in Table 9. As was the case for the atmospheric parameters of the PL scheme, measures of humidity (variables 34, 48, 49, 51) and static stability (variables 38-40, 56) are selected most frequently.

Table 8: List of most frequently selected predictors for PL scheme. The first column refers to No given in Tables 6 and 7; second column gives the percentage a given predictor was selected; the last three column are the average, minimum, and maximum step number of the stepwise regression when it was included.

Predictor	% Selected	Avg Step No	Min Step No	Max Step No
92	73.7	3.1	1	20
82	70.3	13.3	6	20
88	61.7	9.6	3	20
77	61.1	8.0	3	20
76	61.0	10.7	4	20
7	59.8	8.7	1	20
84	59.3	7.7	1	20
21	57.3	8.2	3	20
40	55.4	6.4	1	20
4	52.7	7.1	1	20
90	48.5	7.6	1	20
83	47.9	15.7	8	20
75	44.6	6.4	2	20
32	44.1	11.1	5	20
16	43.5	7.9	1	20
41	42.5	8.7	2	20
74	41.1	10.7	4	20
33	41.1	3.8	1	20
34	40.8	8.2	1	20
93	40.7	17.1	12	20
58	39.3	7.9	1	20
44	37.4	12.9	7	20
71	35.2	9.7	3	20

Table 9: List of most frequently selected predictors for regression of residuals of average cloud cover scheme. See Table 8 for explanation.

Predictor	% Selected	Avg Step No	Min Step No	Max Step No
65	77.2	4.8	1	15
38	71.2	4.3	1	18
56	67.8	4.2	1	17
39	64.2	5.9	1	16
51	54.4	5.2	1	14
40	51.5	3.2	1	15
49	50.1	6.0	1	16
27	49.0	5.9	1	15
64	47.9	5.9	2	14
63	40.9	3.2	1	16
34	38.8	4.3	1	16
48	38.1	5.4	1	18
26	34.7	6.7	2	15
53	34.2	6.6	1	19

### 4.3. Slingo Scheme

The scheme used in the radiative parameterization of the APGSM was included in the comparison of cloud forecast schemes to provide an independent reference point for the verification scores. It is based on the scheme of Slingo (1989; 1987), with some minor modifications introduced by PL to take into account the known biases of the APGSM. Because the radiative parameterization is only invoked every other gridpoint in the APGSM, Slingo cloud scheme forecasts were generated in a postprocessing step, using a modified stand-alone version of the APGSM. No attempt was made to tune the scheme for better agreement with RTNEPH observations.

### 4.4. Persistence

Because clouds are not explicitly predicted by the current generation NWP models, cloud data at the initial time are used, if at all, only in combination with various other observations of moisture, and subsequently interpolated, preprocessed, initialized. After completion of the forecast, the forecast model

output is then converted back to cloud forecasts. Because of errors introduced during these steps, and because of NWP model spinup of the moisture fields, cloud forecasts at the early stages of the NWP forecast are error-prone, making even the most basic forecast (persistence) competitive. For this reason we included persistence as one of the schemes in the cloud forecast verification. Because timely cloud data is not available at all RTNEPH points at any given time, the persistence forecasts were generated from the compacted cloud data of not only the current analysis, but all available previous compacted cloud analyses. Cloud analyses were available starting at the beginning of the each of the four months, resulting in a number of gridpoints with "missing" persistence forecasts, particularly in the first few days of each month.

#### **4.5. GWC 5-layer Forecasts**

The cloud forecasts produced operationally by GWC in the extratropics are produced by the so-called 5-layer model. It is primarily an advective scheme, which uses the GWC GSM predicted winds to advect a moisture variable (condensation pressure deficit, or CPS) deduced from the RTNEPH cloud data. Physical processes such as precipitation are represented by simple parameterizations. It is run (and verified) on the 1/2-mesh octagon. We did not implement and run this scheme because of the prohibitive logistical difficulties involved. However, some verification statistics for the Northern Hemisphere octagon for 1989 were obtained from GWC, and these will be discussed in the next section.

### **5. Results of Cloud Forecast Verification**

Cloud forecasts were generated from the APGSM model output for all forecasts in January (two four-day forecasts each day) and April, July, and October (one four-day forecast every three days). For January, verification statistics were computed for the following cloud forecast schemes:

- Average cloud cover scheme (denoted "avg" in the Figures of verification scores): the forecast cloud amount is the average observed cloud amount of the forecast regime (see section 4.2.2). Total cloud amount is computed from the layer cloud amounts using a tuned overlap.
- CCA scheme ("cca"): Cloud cover is computed from CCA curves of layer-average RH, based on statistics stratified by regime (see section 4.2.3).

Total cloud amount is computed from the layer cloud amounts using a tuned overlap.

PL scheme ("mlr"): Cloud cover (in three terrain-following decks) is computed by multiple linear regression (see section 4.3). Total cloud amount is computed from the layer cloud amounts using a tuned overlap.

Regression of residuals ("rmlr"): Forecasts from the "avg" scheme are refined using a multiple linear regression of its residuals (see section 4.3). Total cloud amount is computed from the layer cloud amounts using a tuned overlap.

Slingo ("sli"): Cloud amount is computed by the Slingo scheme used in the radiative parameterization of the APGSM (see section 4.4).

Persistence ("per"): the forecast cloud amount at any given RTNEPH gridpoint is equal to that of the most recent available cloud data at the initial time of the forecast (see section 4.5).

Because the regression-based schemes ("mlr" and "rmlr") require statistics collected from the forecasts of the previous 10 days, these could not be computed for the months of April, July, or October.

We concentrated on the verification of total cloud cover, because it is the most reliable quantity of the RTNEPH data, and because one of the schemes ("mlr") uses a different set of cloud layers than the others. A number of verification scores were computed, including:

bias: mean forecast error

rmse: root mean square error

correlation: correlation between forecast and verifying cloud cover

sharpness: percentage of forecasts (or observations) that are within 20% of either clear or overcast (i.e., <20% or >80%).

These statistics were computed over the RTNEPH grid octagon, and areas of the RTNEPH grid encompassing the North American and Eurasian continents (see Figure 49). The two land areas were chosen to more closely correspond to the verification regions used by GWC ("fronthalf" and "backhalf"; Kiess, personal communication, 1994); however, because exact information on the gridpoint boundaries of these was unavailable, the correspondence is only approximate. Verification statistics were also computed on the GSM transform grid, stratified by the weather regimes. Statistics were computed only for gridpoints with timely RTNEPH data, and with non-missing persistence forecasts (to ensure a fair comparison between the forecast schemes and persistence).

## 5.1. Verification Statistics for January

Verification statistics for all 6 cloud forecast schemes were computed for 12 - 96 hour forecasts starting from January 11 through January 27 (forecasts from the first 10 days in January were used to collect the statistics for the regression schemes). The bias is shown in Figure 58 for the entire NH RTNEPH octagon, and for the fronthalf and backhalf subregions (see Figure 49) in Figures 59 and 60. While the bias in total cloud cover over the entire octagon is small for all techniques, values are slightly larger for the Slingo technique and the CCA technique when computed over just the fronthalf or the backhalf. Layer cloud cover values show a positive bias of the Slingo scheme at the 300 hPa layer (and, to a lesser extent, the 400 hPa layer). Averaged over the entire octagon, the Slingo scheme has a negative bias at lower layers, which is largely absent in the continental subregions. The bias of all the other schemes is small (less than 10% at most layers and regions), as would be expected since both the regime-based (avg and cca) and regression-based (mlr and rmlr) schemes are designed to predict the correct mean value. The rms errors of cloud cover are shown in Figures 61 - 63. In general, rms errors are much larger than the bias, resulting in very similar curves for the rmse and error standard deviation (not shown). In terms of total cloud cover, the Slingo scheme has the largest errors by far (slightly above 50% at all forecast lead times). Of all the schemes, persistence is the only one showing an appreciable dependence of error magnitude on forecast lead time: its errors increase by approximately 10% over the first 36 hours of the forecast. As a result, it is competitive with all schemes at the beginning of the forecast, but by 36 hours it is worse than all but the Slingo scheme. As was the case with the dependent data set (section 2.4.3), the CCA scheme has larger rms errors than the average scheme. In terms of rms errors, it is slightly better than persistence after 36 hours (except over the backhalf region, where the two are essentially equal). Both regression schemes have rms errors of about the same magnitude as the average scheme. Statistics for the layer cloud cover generally follow the same pattern as those for total cloud cover (but recall that layer cloud cover statistics are not computed for the mlr scheme). In particular, the rmlr and avg scheme are virtually indistinguishable in terms of rmse, indicating that the regression of the residuals provides almost no additional information. The correlation scores shown in Figures 64 - 66 exhibit larger differences between the techniques, with highest scores for the mlr, followed by the rmlr, avg, and cca

schemes. Slingo is consistently worst, and persistence varies widely with region and layer. A very different ranking emerges when one considers the sharpness of the cloud forecasts (Figures 67 - 69). Here, the Slingo scheme is closest to the perfect value of persistence, although it produces cloud distributions that are too binary. The CCA scheme forecasts of layer cloud cover are slightly too smooth, resulting in a total cloud cover sharpness of less than 40% (compared to the verifying value of near 65%). The avg, mlr, and rmlr schemes all produce forecasts that are much too smooth, with total cloud cover sharpness values of 10% or less. The same holds true for the layer values, except at the upper layers, where mean values are below 20% in many cases. Thus, higher rmse and correlation skill scores of the avg, mlr, and rmlr schemes come at the expense of a highly unrealistic frequency distribution of cloud cover.

The verification statistics of the non-regression schemes were also computed for all forecasts in January (including the first 10 days). The results (not shown) showed no appreciable differences from those for January 11-27.

## **5.2. Verification Statistics for April, July, and October**

The seasonal dependence of the results shown for January can be seen in the corresponding statistics computed for the months of April, July, and October. Curves for the bias (for the entire RTNEPH octagon only) are shown in Figures 70 - 72. The Slingo scheme has a consistently negative bias in total and low level cloud cover in all four months, and its magnitude is substantially larger than that of all the other schemes in July and October. The corresponding curves for the rms errors are shown in Figures 73 - 75. In all three months, persistence forecasts of total cloud cover have larger rms errors than all but the Slingo scheme after the first 12-24 hours of the forecast. This holds true not only when computed over the entire octagon, but also separately over the fronthalf and backhalf regions (not shown). As was the case in January (Figure 61), Slingo has the largest, and the average scheme the lowest errors. The ranking of the average, CCA, and Slingo schemes in terms of correlation scores (Figures 76 - 78, and Figure 64) is the same as in terms of the rmse scores, but persistence scores vary more widely with level and month. Finally, the sharpness scores shown in Figures 79 - 81 show virtually no seasonal change for total cloud cover. At individual layers, the average scheme shows some variability, reflecting seasonal changes in the frequency of certain regimes.



### 5.3. Comparison with GWC Verification Statistics

The GWC cloud forecast verification statistics available to us were 5-layer correlation, rmse, and error standard deviation scores for the fronthalf and backhalf, at forecast lead times of 12, 24, 36 and 48 hours, and NH octagon bias and error standard deviation scores for 5-layer and persistence forecasts at 24 hours. Because of possible differences in the fronthalf and backhalf regions used here and by GWC, we concentrated on the octagon verification statistics. Table 10 shows a comparison of rms errors of total cloud covers computed by GWC (shown for both the 5layer and persistence), and those computed here. For the month of January, verification statistics are shown both for the whole month and for the part for which the regression scheme forecasts were available. It can be seen that both the forecast and persistence rms errors computed by GWC are smaller, possibly due to differences in the verification procedure (including the compaction of the verifying RTNEPH data) and sampling differences. In terms of a percent improvement over persistence, the average scheme is consistently better than the 5-layer model, whereas the CCA scheme does consistently worse. The two regression schemes have smaller rms errors, both in absolute and relative (to persistence) terms than any of the other schemes. Thus, at 24 hours, only the schemes with very smooth cloudiness distributions improve on the currently operational 5-layer scheme in terms of rms errors.

Table 10: Comparison of 24 hour total cloud cover rmse scores computed over the NH octagon. For each method, values are shown for the forecast rmse (Fcst), persistence rmse (Per), and the percentage improvement over persistence (Skill).

Method	Jan			Apr			Jul			Oct		
	Fcst	Per	Skill	Fcst	Per	Skill	Fcst	Per	Skill	Fcst	Per	Skill
5layer	33.95	34.49	1.5	33.90	36.70	7.6	32.82	35.44	7.3	32.80	35.90	8.6
CCA	38.04	36.62	-3.0	36.89	39.20	5.8	37.99	39.16	2.9	36.54	39.76	8.0
CCA	38.09	37.08	-2.7	part of January for which regression schemes were used								
Avg	34.61	36.62	5.0	33.64	39.20	14.1	35.32	39.16	9.8	34.11	39.76	14.2
Avg	34.62	37.08	6.6	part of January for which regression schemes were used								
Sli	52.20	36.62	-42.0	47.63	39.20	-21.5	49.14	39.16	-25.4	48.97	39.76	-23.1
Sli	52.39	37.08	-41.2	part of January for which regression schemes were used								
mlr	32.50	37.08	12.3	part of January for which regression schemes were used								
rmlr	33.04	37.08	10.8	part of January for which regression schemes were used								

Since the 5-layer model is primarily an advective model, one might expect larger errors at longer forecast lead times, as a result of the lack or misrepresentation of the moist physics. An indication of the cloud forecast error growth of the 5-layer model is given in Table 11, which shows the fronthalf and backhalf correlation and rmse scores for the first 48 hours of the forecast for January. Although a direct comparison of the absolute score values with those of Figures 62, 63, 65, and 66 may be inappropriate (because of differences in verification procedures), it appears that the verification scores are initially much better than, but approach the values of the CCA scheme by 48 hours.

Table 11: Evolution of GWC 5-layer total cloud cover correlation and rmse scores with forecast lead time (12-48 hours), for January, April, July, and October 1989, for the fronthalf (-f) and backhalf (-b).

Month /Area	Correlation				RMSE			
	12-hr	24-hr	36-hr	48-hr	12-hr	24-hr	36-hr	48-hr
Jan-f	0.64	0.58	0.46	0.41	31.0	34.4	39.3	41.4
Jan-b	0.68	0.60	0.50	0.45	29.9	33.7	38.1	40.2
Apr-f	0.57	0.53	0.39	0.36	32.1	34.1	39.2	40.5
Apr-b	0.65	0.58	0.47	0.42	30.5	33.8	38.5	40.3
Jul-f	0.53	0.50	0.36	0.35	31.9	34.0	38.3	39.2
Jul-b	0.66	0.65	0.55	0.54	30.6	31.5	35.7	36.2
Oct-f	0.63	0.56	0.43	0.39	30.1	33.8	38.5	40.5
Oct-b	0.72	0.67	0.58	0.54	27.6	31.0	34.7	36.7

## 6. Summary and Conclusions

This report describes the results of a 30-month study designed to demonstrate that the advanced weather prediction model developed at the Geophysics Directorate of the Phillips Laboratory (the APGSM) is capable of fulfilling the forecasting needs of the US Air Force. The currently used GWC GSM, obtained from the National Meteorological Center (NMC) as far back as 1984, has a very simple set of physical parameterizations. Forecasts of conventional variables (temperature, geopotential height, winds, and RH) from both models (GWC GSM and APGSM) were compared in side-by-side tests, using data from all four seasons of 1989. The operational GWC analyses (High Resolution Analysis System, or HIRAS) were used for initializing and verifying the forecast models.

The APGSM shows only one area in which it consistently improves on the performance of the GWC GSM, namely the warm bias of the extratropical troposphere (in the tropics, it has a larger warm bias than the GWC GSM in the upper troposphere). In practically every other measure of skill, APGSM results are worse than those of the GWC GSM during the first 1-2 days of the forecast. At later forecast lead times, results are mixed. In the extratropics, the APGSM has smaller RMS errors of geopotential at the upper levels, and of RH at 1000 and 850 hPa. RMS errors of the horizontal wind (vector errors and u and v components) are generally larger for the APGSM, as are the RH bias and RMS

errors at upper levels. In general, the degradation of the APGSM performance is worse in the tropics. Possible reasons for the relatively poor performance of the APGSM are deficiencies in the model initialization procedure (since model errors are degraded particularly at the beginning of the forecast), in the gravity wave drag or PBL formulation and/or the orography data set (since wind errors are worsened more than those of geopotential height), and in the convection scheme, and/or its interaction with the radiation scheme (since large biases were found in the tropics). No direct evidence was found to indicate that the use of HIRAS analyses for verification unfairly favored the GWC GSM.

GWC operational cloud forecasts are based on trajectory models with highly parameterized physics, and utilize only the wind fields of the forecast model output. As a potential alternative, schemes were developed and evaluated that use the APGSM model output to generate cloud forecasts. The cloud forecasts were verified against RTNEPH cloud analyses. Schemes developed under this contract use a classification of vertical profiles of atmospheric variables into weather regimes. Either the mean cloud cover of each regime (avg scheme), or a cloud-RH curve derived for each regime (cca scheme), were used to generate cloud forecasts. These were compared against forecasts from a multiple linear regression scheme (mlr) developed by PL, the Slingo scheme used in the APGSM, and persistence. An additional scheme (rmlr) used a regression equation for the residuals of the regime-average scheme.

In terms of rms errors of total cloud cover, the Slingo scheme has the largest errors of all the schemes. It must be noted that it is the only scheme tested that was not tuned with or derived from the RTNEPH data used for verification in this study, so that this result does not indicate that this scheme will also be inferior after careful tuning. Persistence has as low or lower errors as the other schemes at the beginning of the forecast (12 hours), but by 36 hours it is inferior to all but the Slingo scheme. CCA errors are larger than the avg, mlr, and rmlr errors, which are all very close. The correlation scores exhibit larger differences between the techniques, with highest scores for the mlr, followed by the rmlr, avg, and cca schemes. Slingo is consistently worst, and persistence varies widely with region and layer. In terms of the sharpness of the forecast, however, the CCA (and Slingo) scheme is more realistic than the avg, mlr, or rmlr schemes, which are all much too smooth. Thus, higher rmse and correlation skill scores of

the avg, mlr, and rmlr schemes come at the expense of a highly unrealistic frequency distribution of cloud cover. The somewhat disappointing results, which indicate that only a small part of the variance of the observed cloud cover can be resolved by any of the schemes, point to the limiting factors of this study: the accuracy of the APGSM forecasts and RTNEPH cloud analyses, and the scatter inherent in any simple relationship between atmospheric parameters and cloud cover.

Comparison with verification statistics computed by GWC for the operational suite of cloud forecast models was limited by the data availability and differences in verification procedure. Indications from the limited sample are that by 24 hours, rms scores (measured relative to persistence) of the current 5-layer model fall somewhere in between the cca and avg, mlr, and rmlr schemes. By 48 hours, errors of the 5-layer scheme approach those of the CCA scheme.

## **7. References**

- Ballish, B., 1980: Initialization, theory and application to the NMC spectral model. Ph.D. Thesis, Department of Meteorology, University of Maryland
- Brenner, S., C. H. Yang and S. Y. K. Yee, 1982: The AFGL spectral model of the moist global atmosphere. AFGL-TR-82-0393. Air Force Geophysics Laboratory, Hanscom AFB, MA. 65 pp. [NTIS ADA129283]
- Hamill, T. M., R. P. d'Entremont and J. T. Bunting, 1992: A description of the Air Force Real-Time Nephanalysis model. *Wea. Forecasting.*, **7**, 288-306.
- Kuo, H. L., 1965: On formation and intensification of tropical cyclones through latent heat release by cumulus convection. *J. Atmos. Sci.*, **22**, 40-63.
- Liou, K.-N., S.-C. Ou, S. Kinne and G. Koenig, 1984: Radiation parameterization program for use in general circulation models. AFGL-TR-84-0217. Air Force Geophysics Laboratory, Hanscom AFB, MA. [NTIS ADA148015]
- Louis, J.-F., R. N. Hoffman, T. Nehrkorn and D. Norquist, 1989: Observing systems experiments using the AFGL four-dimensional data assimilation system. *Mon. Wea. Rev.*, **117**, 2186-2203.

- Machenhauer, B., 1977: On the dynamics of gravity oscillations in a shallow water model with applications to normal mode initialization. *Beitr. Phys. Atmos.*, **50**, 253-271.
- Mahrt, L., H.-L. Pan, J. Paumier and I. Troen, 1984: A boundary layer parameterization for a general circulation model. AFGL-TR-84-0063. Air Force Geophysics Laboratory, Hanscom AFB, MA. [NTIS ADA144224]
- Mitchell, K. E. and D. C. Hahn, 1989: Development of a cloud forecast scheme for the GL baseline global spectral model. GL-TR-89-0343. Geophysics Laboratory, Hanscom AFB, MA. 147 pp. [NTIS ADA231595]
- Murphy, A. H. and E. S. Epstein, 1989: Skill scores and correlation coefficients in model verification. *Mon. Wea. Rev.*, **117**, 572-581.
- Nehrkorn, T., R. Hoffman, J.-F. Louis and M. Zivkovic, 1992: An enhanced global spectral model. PL-TR-92-2011. Geophysics Directorate, Phillips Laboratory, Hanscom AFB, MA 01731. 52 pp. [NTIS ADA251242]
- Nehrkorn, T., R. N. Hoffman, J.-F. Louis, R. G. Isaacs and J.-L. Moncet, 1993a: Analysis and forecast improvements from simulated satellite water vapor profiles and rainfall using a global data assimilation system. *Mon. Wea. Rev.*, **121**, 2727-2739.
- Nehrkorn, T., M. Mickelson, L. W. Knowlton and M. Zivkovic, 1993b: A global forecast model comparison study - year 1. PL-TR-93-2203. Phillips Laboratory, Hanscom Air Force Base, MA. 40 pp. ADA275674
- Norquist, D. C., H. S. Muench, D. L. Aiken and D. C. Hahn, 1994: Diagnosing cloudiness from global numerical weather prediction model forecasts. Phillips Laboratory, Hanscom AFB, MA. PL-TR-94-2211
- Norquist, D. C., C.-H. Yang, S. Chang and D. C. Hahn, 1992: Phillips Laboratory global spectral numerical weather prediction model. PL-TR-92-2225. Phillips Laboratory, Hanscom AFB, MA. 154 pp. [NTIS ADA 267293]

- Ou, S.-C. and K.-N. Liou, 1988: Development of radiation and cloud parameterization programs for AFGL Global Models. AFGL-TR-88-0018. Air Force Geophysics Laboratory, Hanscom AFB, MA. 88 pp. [NTIS ADA193369]
- Schattel, J., 1992: Refinement and testing of the radiative transfer parameterization in the PL global spectral model. PL-TR-92-2169. Phillips Laboratory, Hanscom AFB, MA. [NTIS ADA 256840]
- Sela, J. G., 1980: Spectral modeling at the National Meteorological Center. *Mon. Wea. Rev.*, **108**, 1279-1292.
- Slingo, A., 1989: A GCM parameterization for the shortwave radiative properties of water clouds. *J. Atmos. Sci.*, **46**, 1419-1427.
- Slingo, J. M., 1987: The development and verification of a cloud prediction scheme for the ECMWF model. *Quart. J. R. Meteor. Soc.*, **113**, 899-928.
- Tiedtke, M., 1989: A comprehensive mass flux scheme for cumulus parameterization in large-scale models. *Mon. Wea. Rev.*, **117**, 1779-1800.
- Vernekar, A. D., J. Zhou and B. Kirtman, 1991: A comparison of systematic errors in AFGL and COLA forecast models. PL-TR-91-2164. Phillips Laboratory, Hanscom AFB, MA. [NTIS ADA244458]
- Yang, C.-H., K. Mitchell, D. Norquist and S. Yee, 1989: Diagnostics for and evaluations of new physical parameterization schemes for global NWP models. GL-TR-89-0158. Geophysics Laboratory, Hanscom AFB, MA. [NTIS ADA228033]
- Zivkovic, M. and J.-F. Louis, 1992: A new method for developing cloud specification schemes in general circulation models. *Mon. Wea. Rev.*, **120**, 2928-2941.

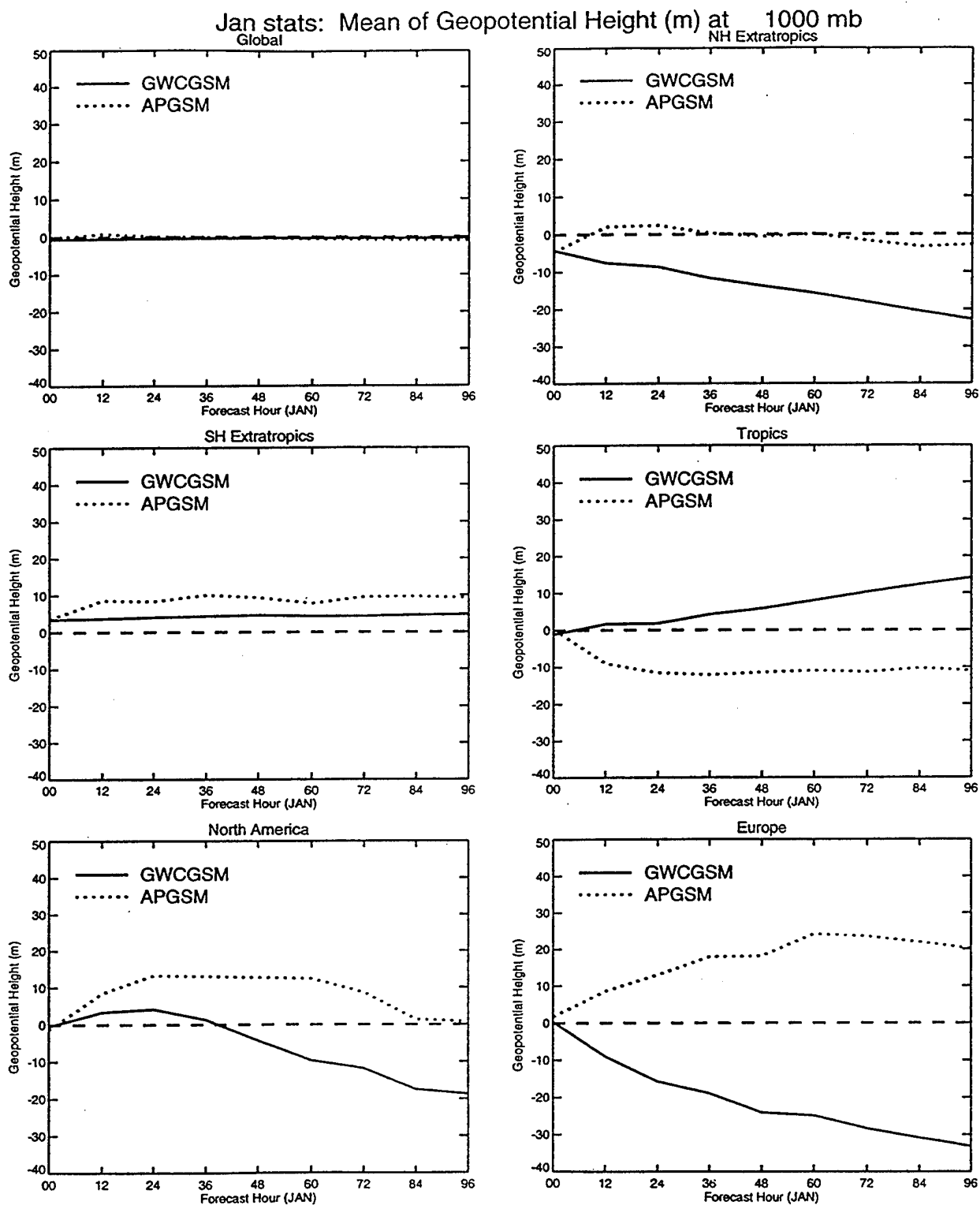


Figure 2: Mean forecast error (bias) of 1000 hPa geopotential height for January forecasts, for the GWC GSM and APGSM. Errors are shown for the geographical regions of Table 1.



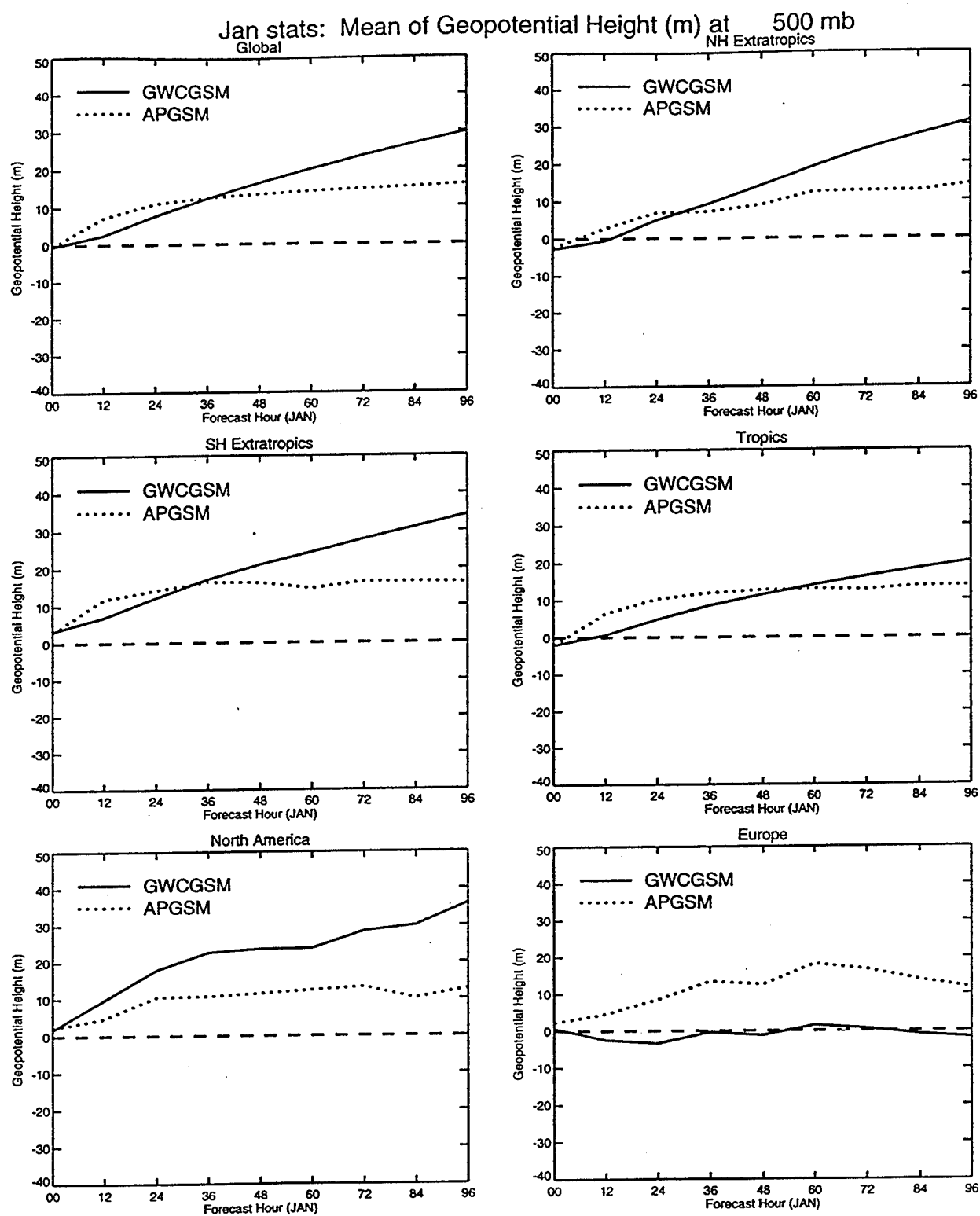


Figure 3: As Figure 2, except for the bias of 500 hPa geopotential height.

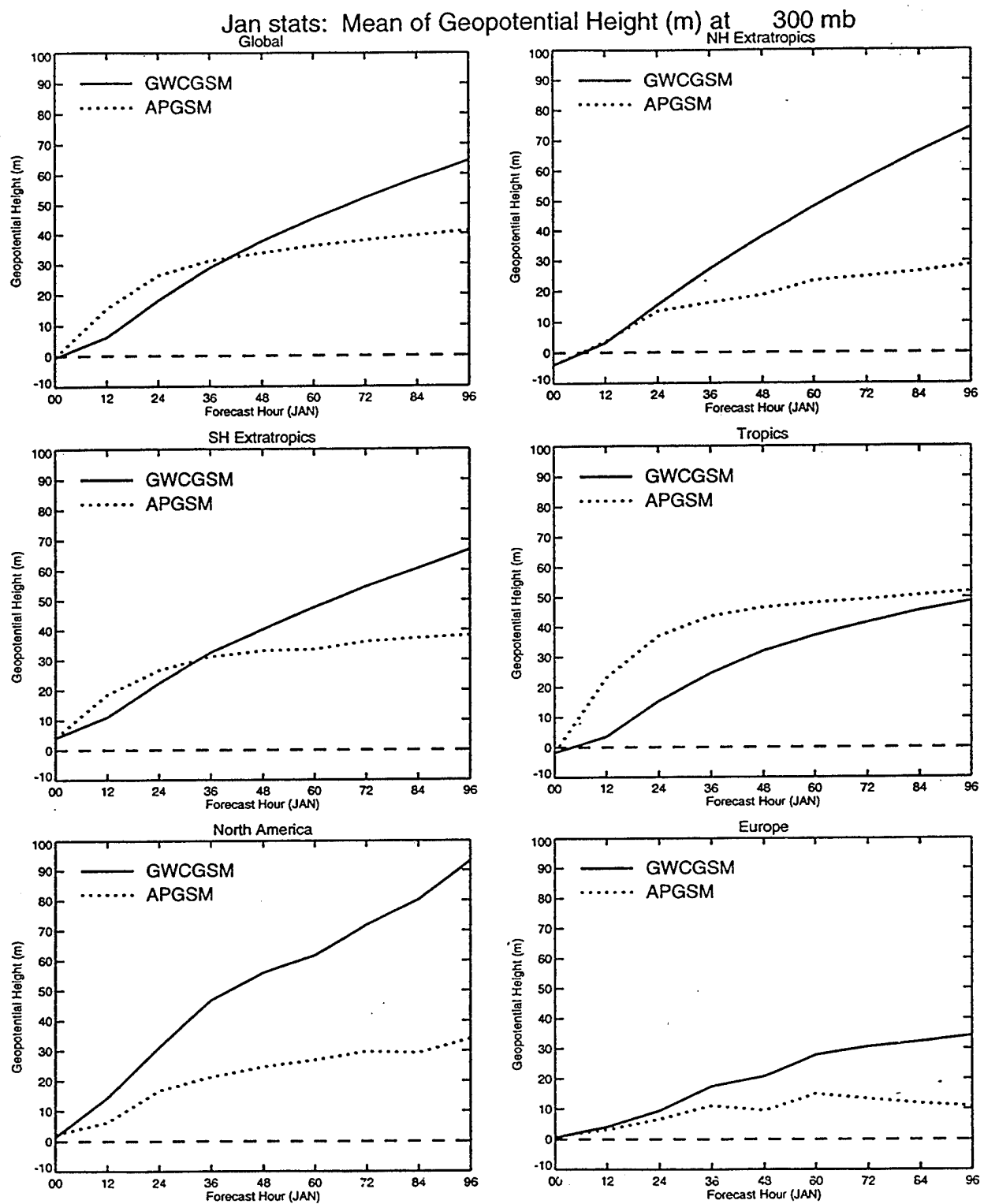


Figure 4: As Figure 2, except for the bias of 300 hPa geopotential height.

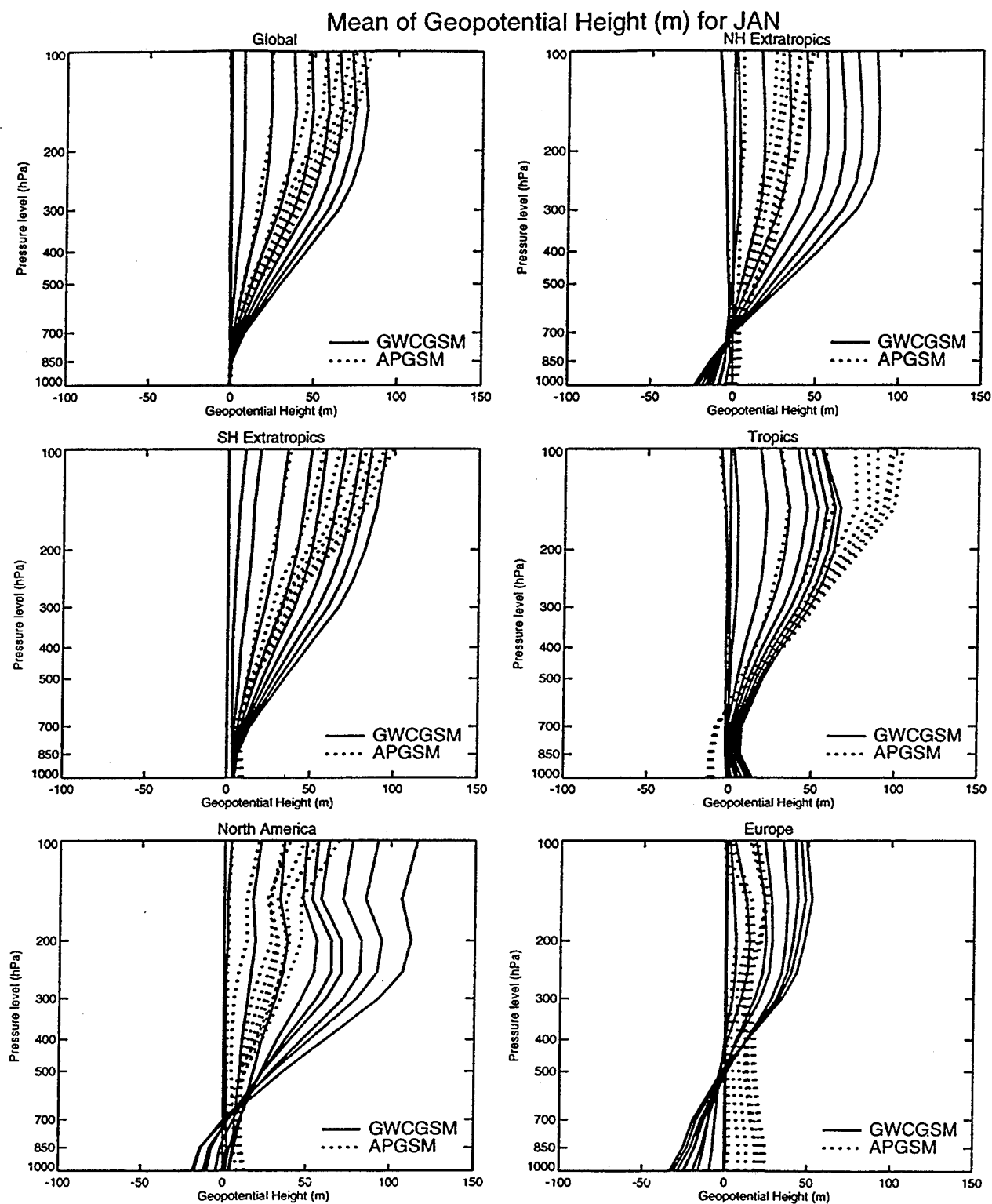


Figure 5: Vertical profile of mean errors of geopotential height for the APGSM (dotted line) and GWC GSM (solid line) for 0, 12, 24, ..., 96 hour forecasts, for the geographical regions of Table 1.

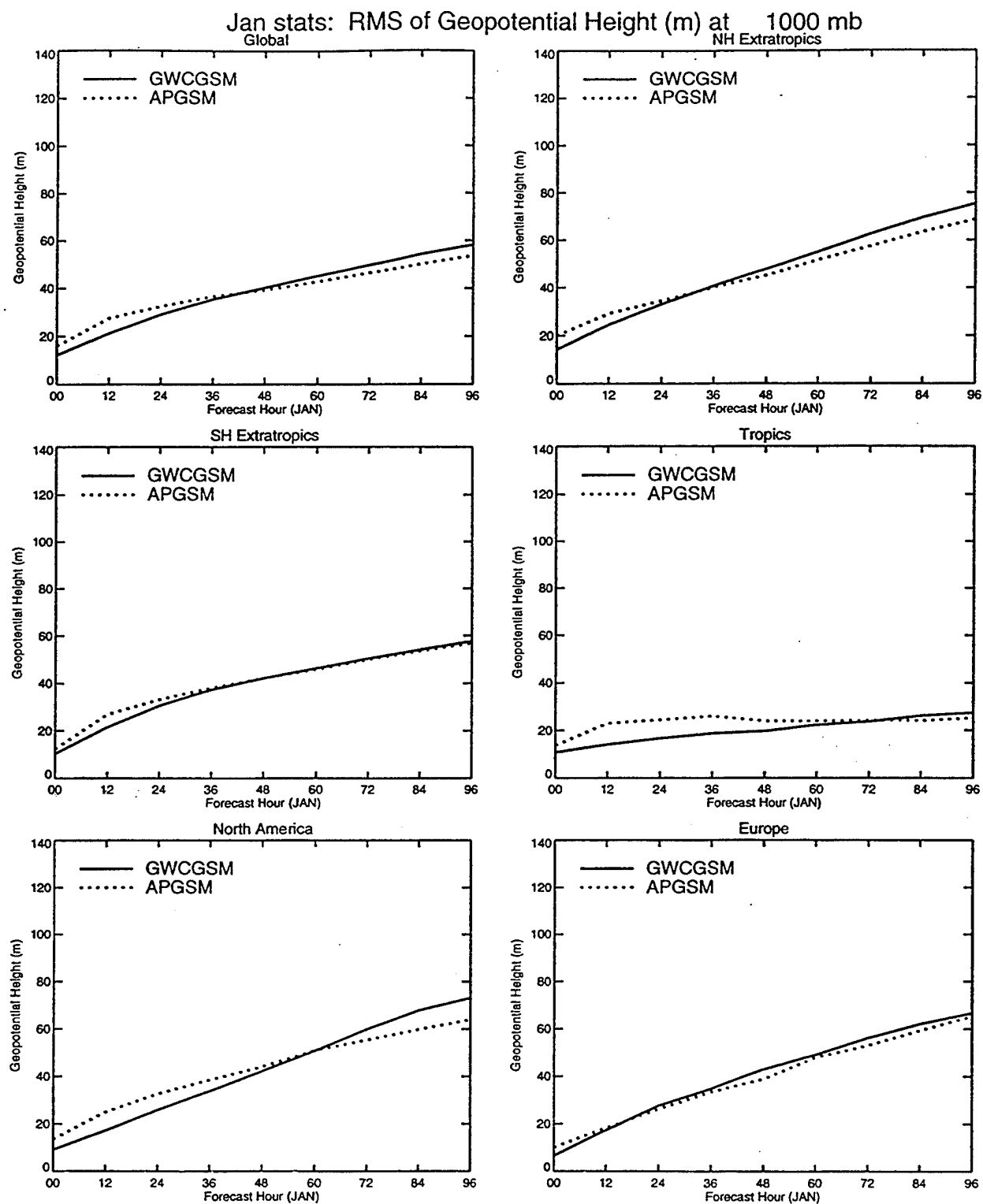


Figure 6: As Figure 2, except for the root mean square (RMS) error of 1000 hPa geopotential height.

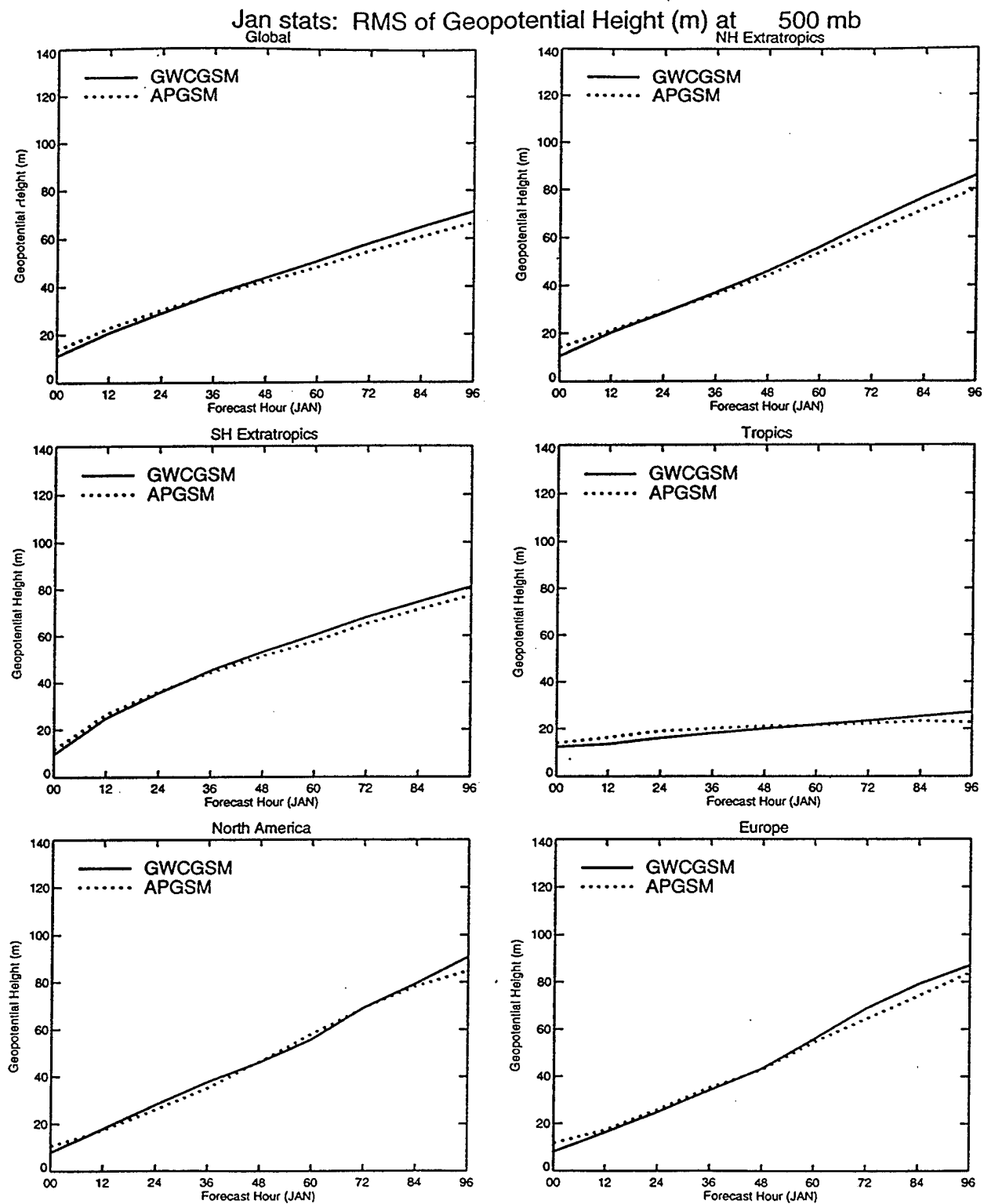


Figure 7: As Figure 2, except for the RMS error of 500 hPa geopotential height.

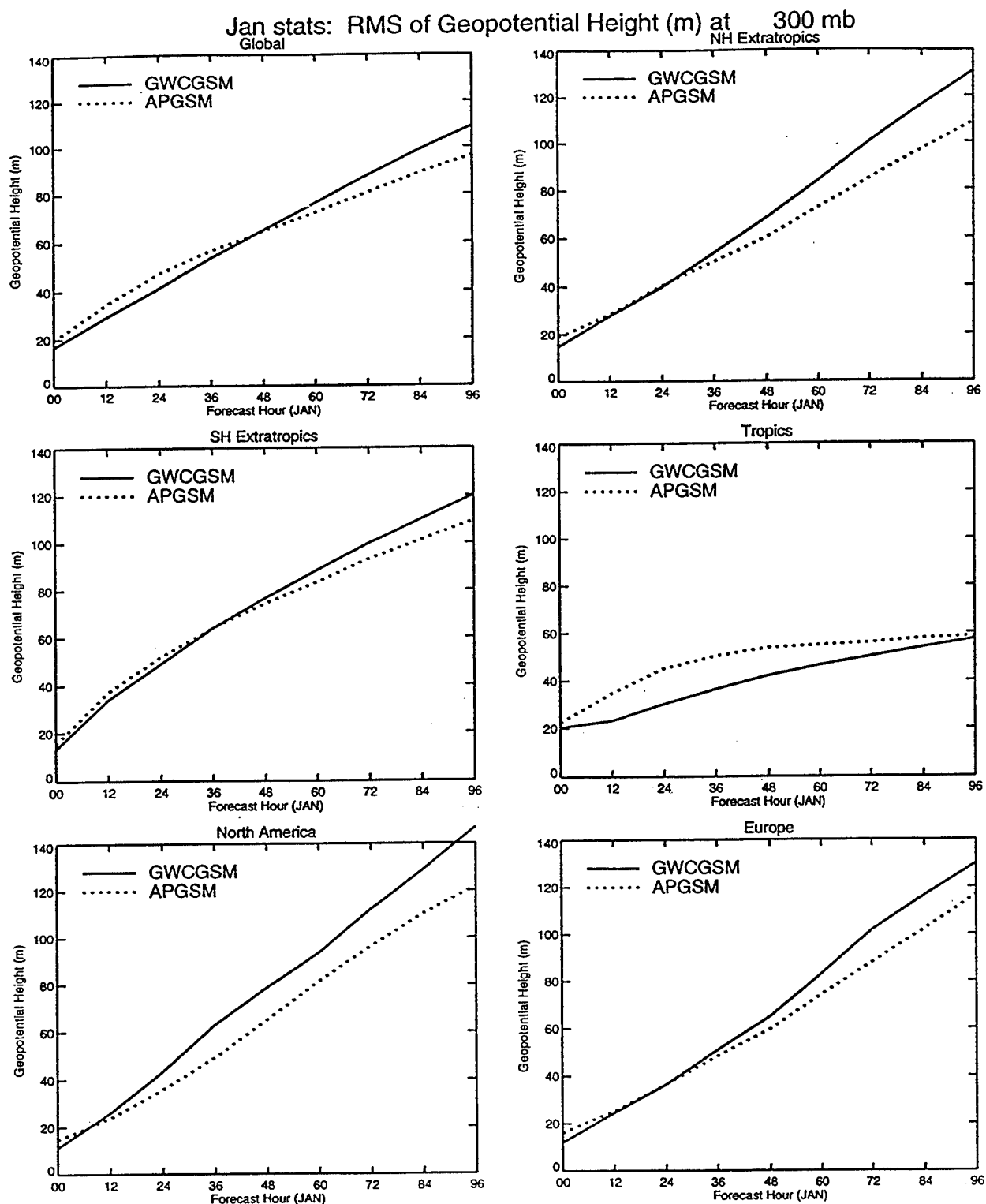


Figure 8: As Figure 2, except for the RMS error of 300 hPa geopotential height.

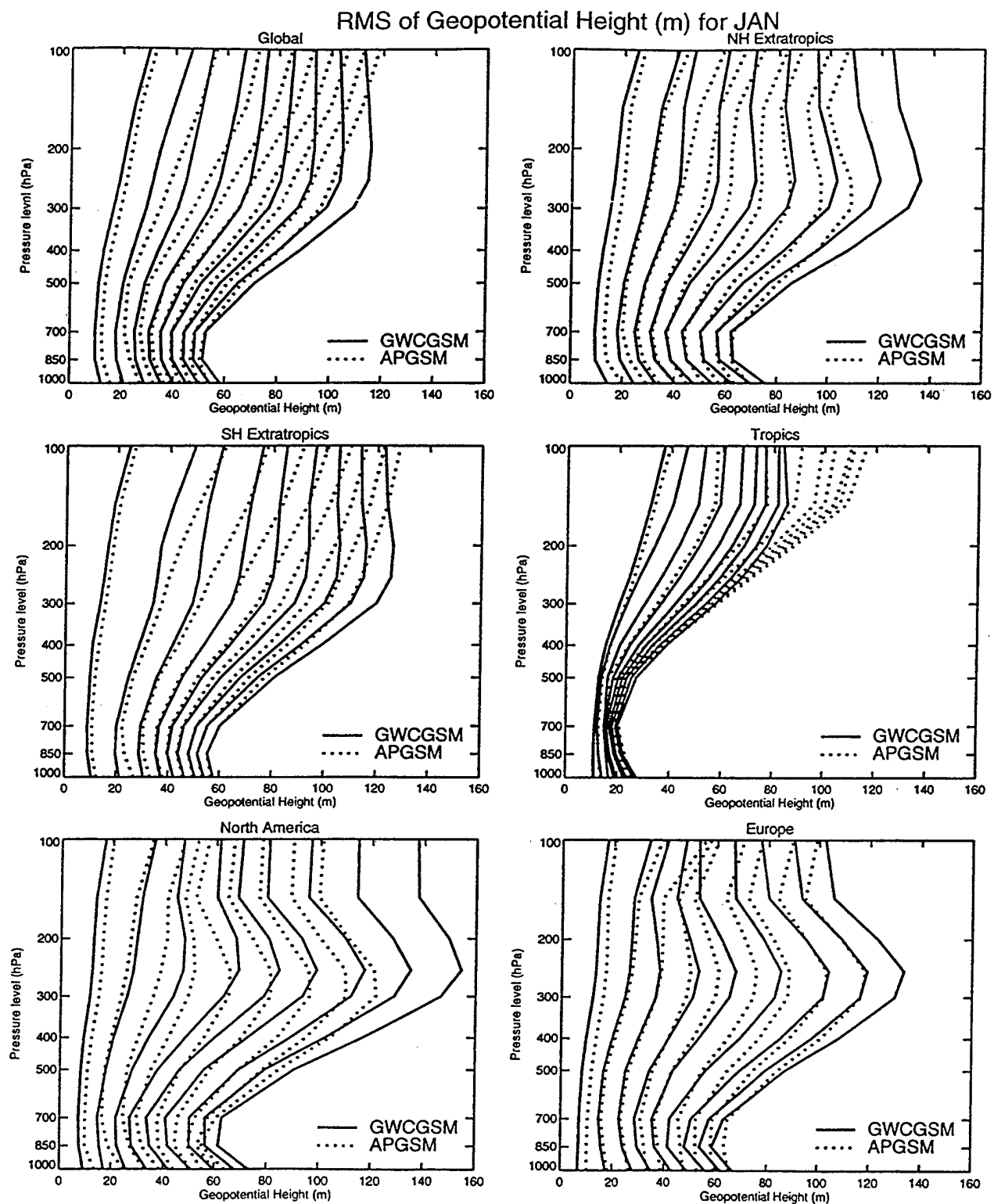


Figure 9: As Figure 5, except for the RMS error of geopotential height.

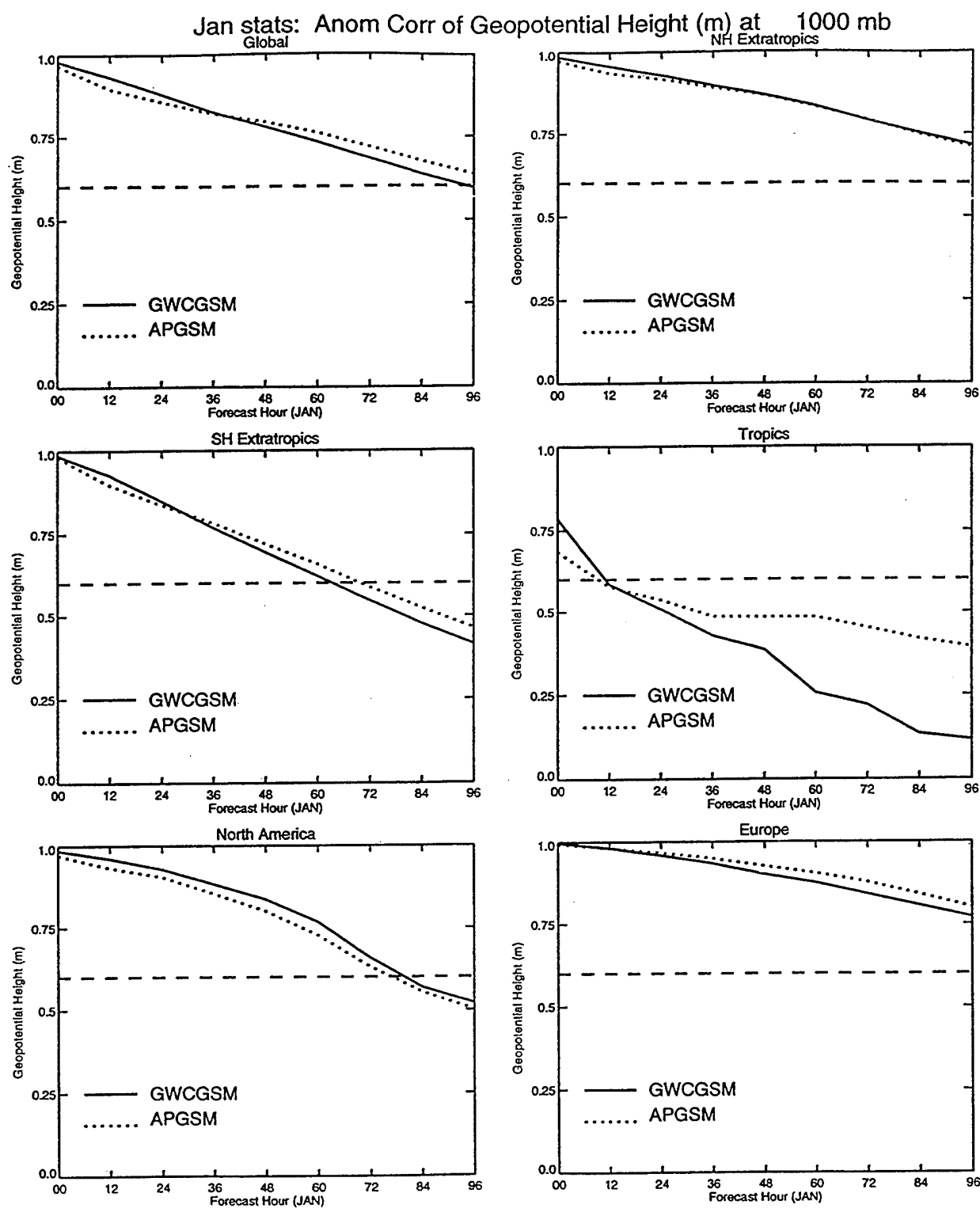


Figure 10: As Figure 2, except for the anomaly correlation of 1000 hPa geopotential height.



# Jan stats: Clim Skill of Geopotential Height (m) at 1000 mb

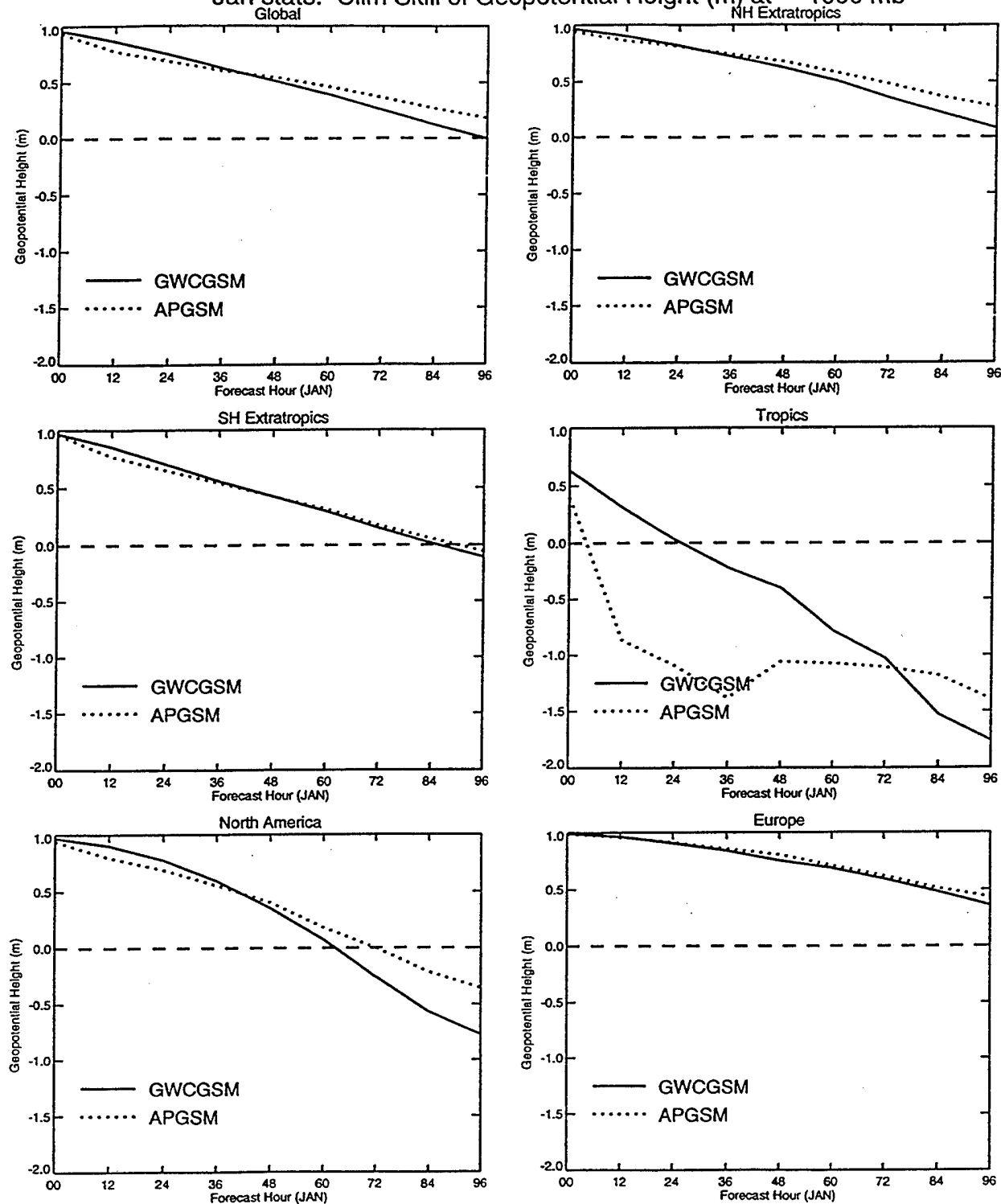


Figure 11: As Figure 2, except for the climatological skill score of 1000 hPa geopotential height.

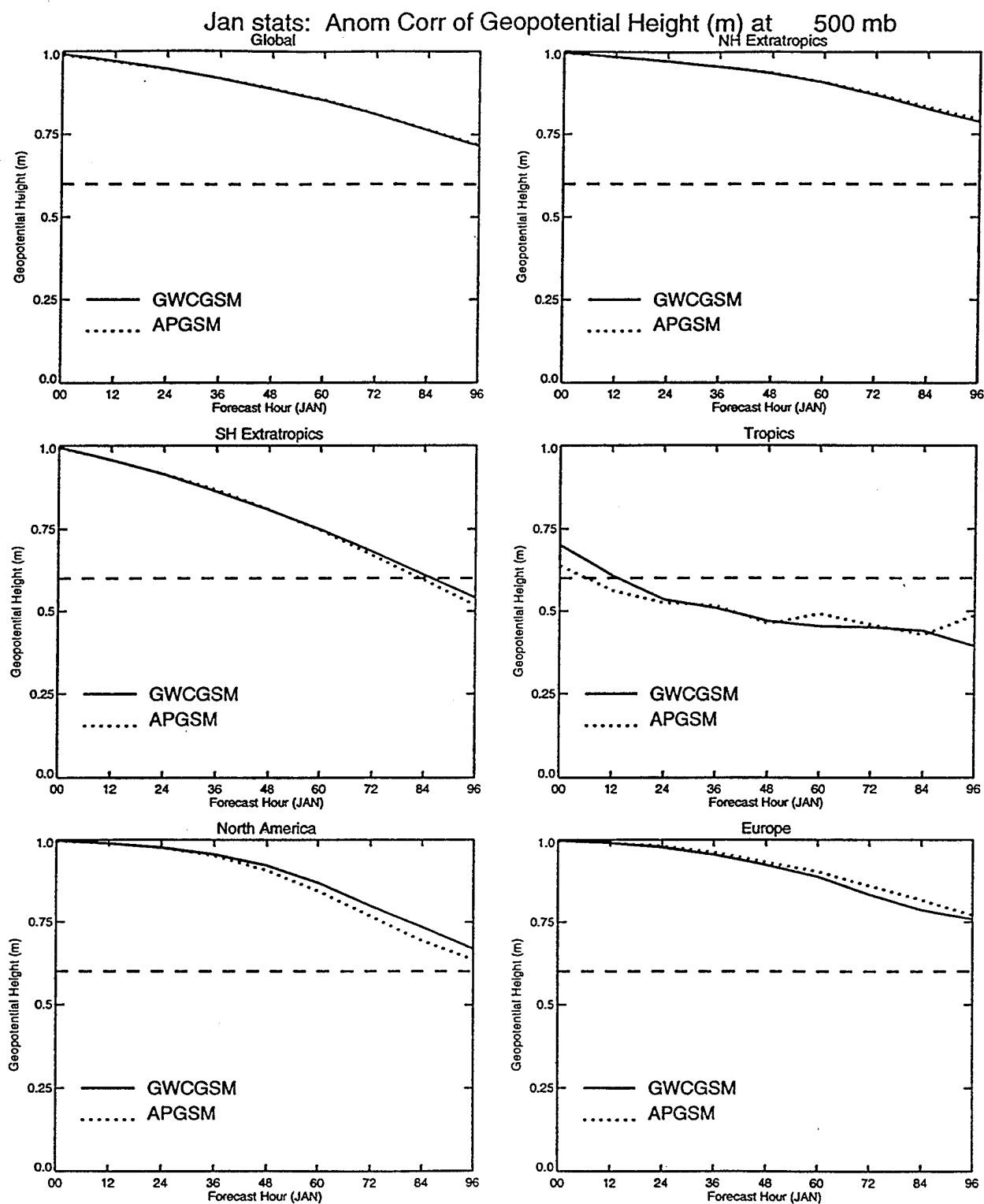


Figure 12: As Figure 2, except for the anomaly correlation of 500 hPa geopotential height.

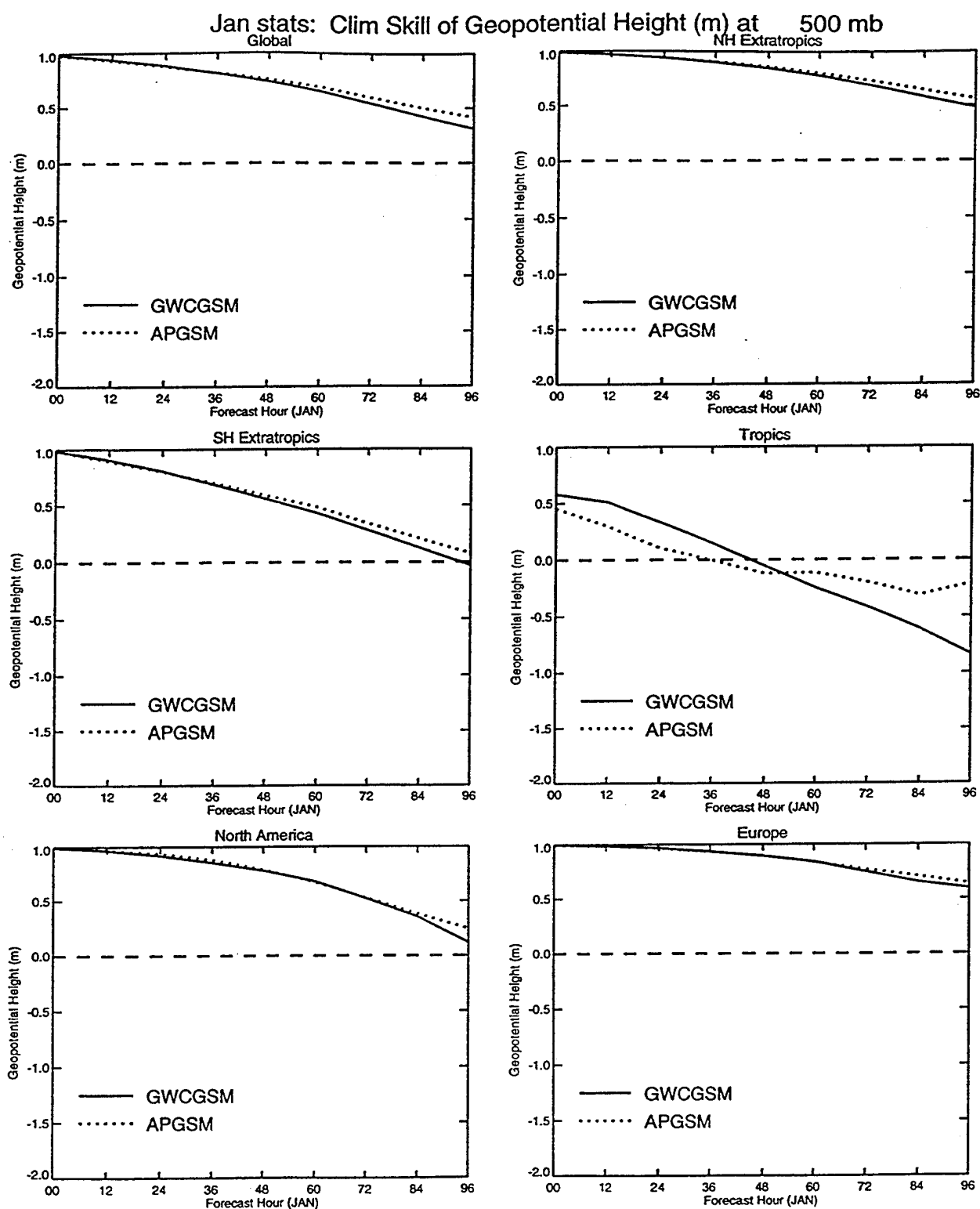


Figure 13: As Figure 2, except for the climatological skill score of 500 hPa geopotential height.

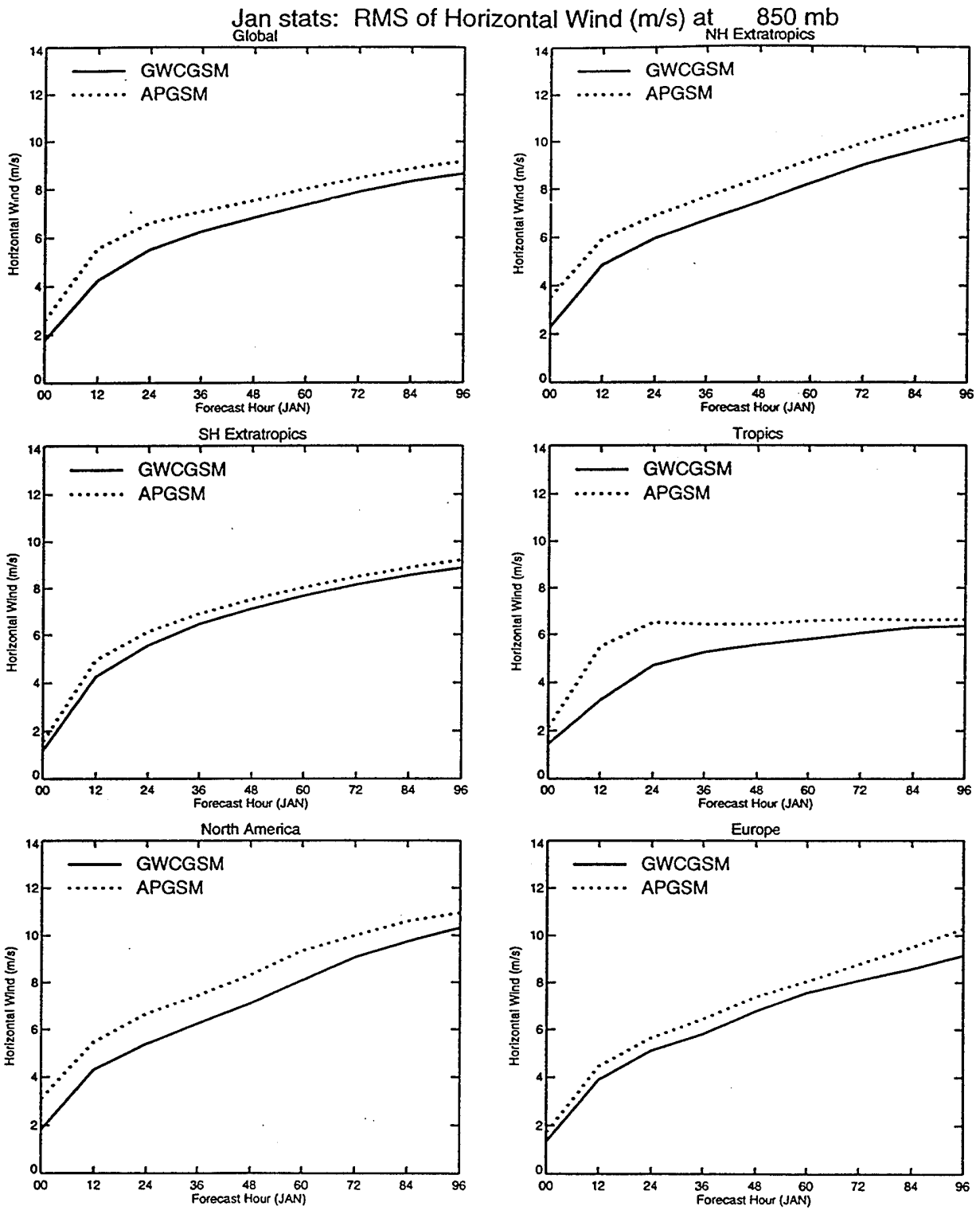


Figure 14: As Figure 2, except for the RMS error of the horizontal wind at 850 hPa.

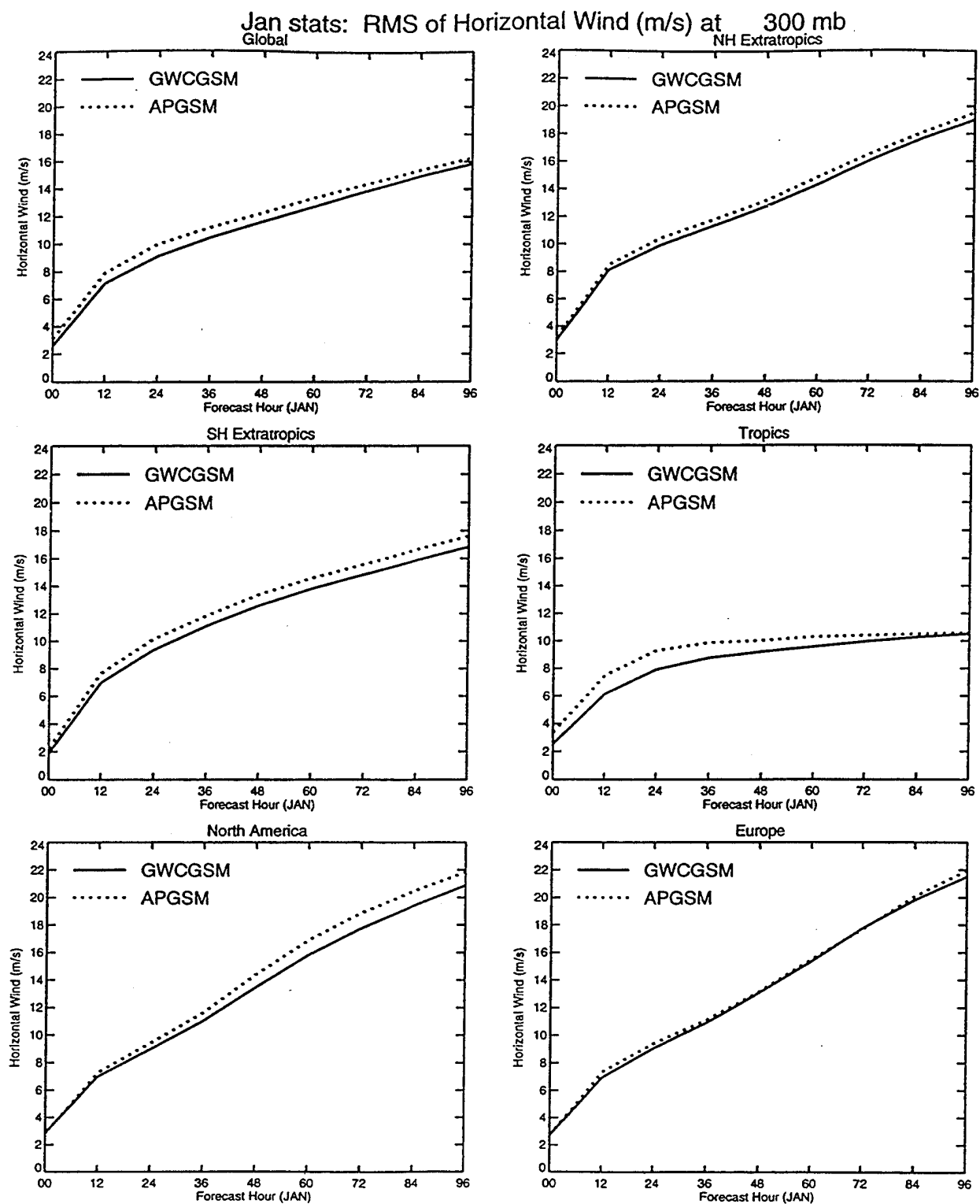


Figure 15: As Figure 2, except for the RMS error of the horizontal wind at 300 hPa.

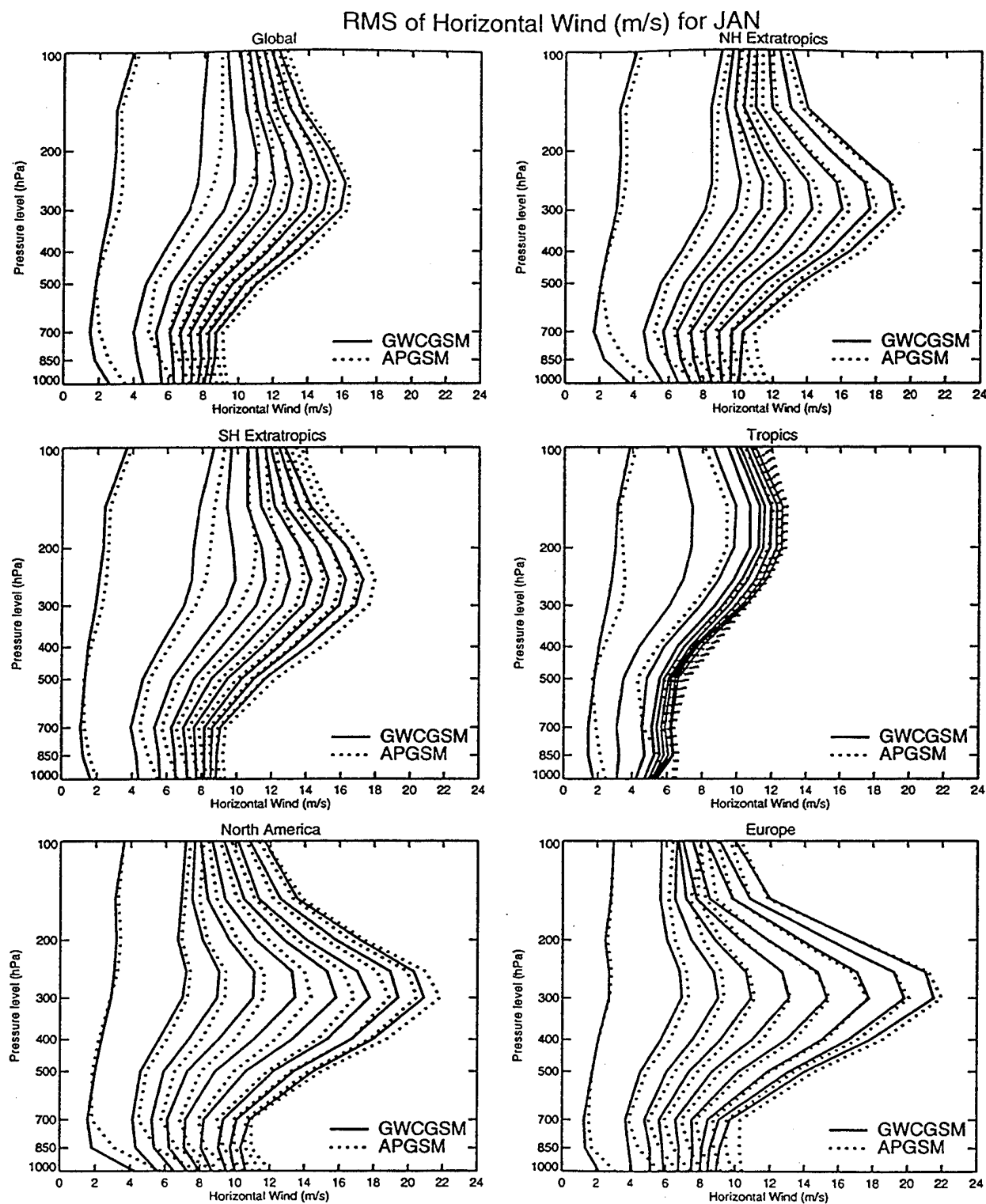


Figure 16: As Figure 5, except for the RMS errors of the horizontal wind.

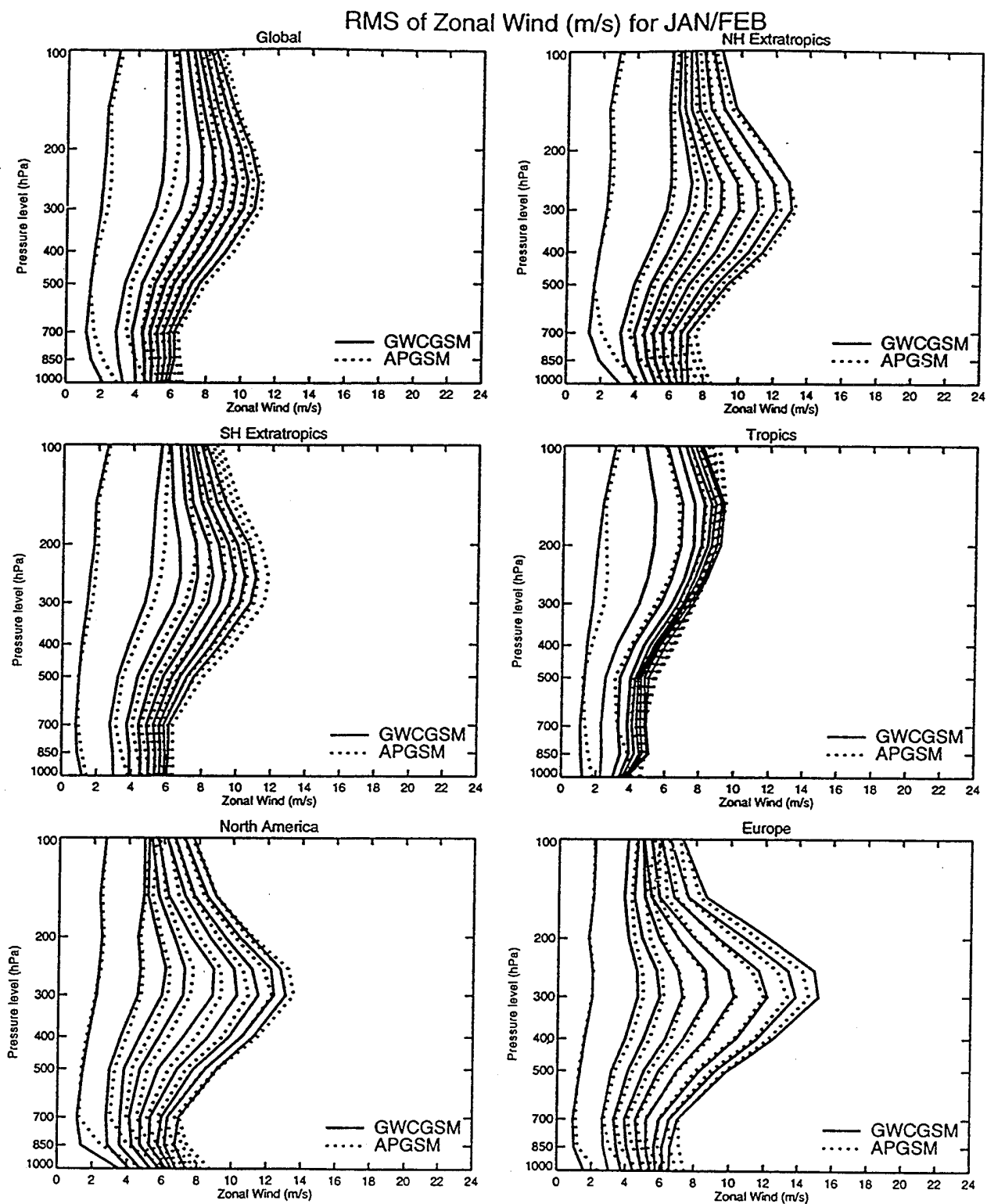


Figure 17: As Figure 5, except for the RMS errors of the zonal wind.

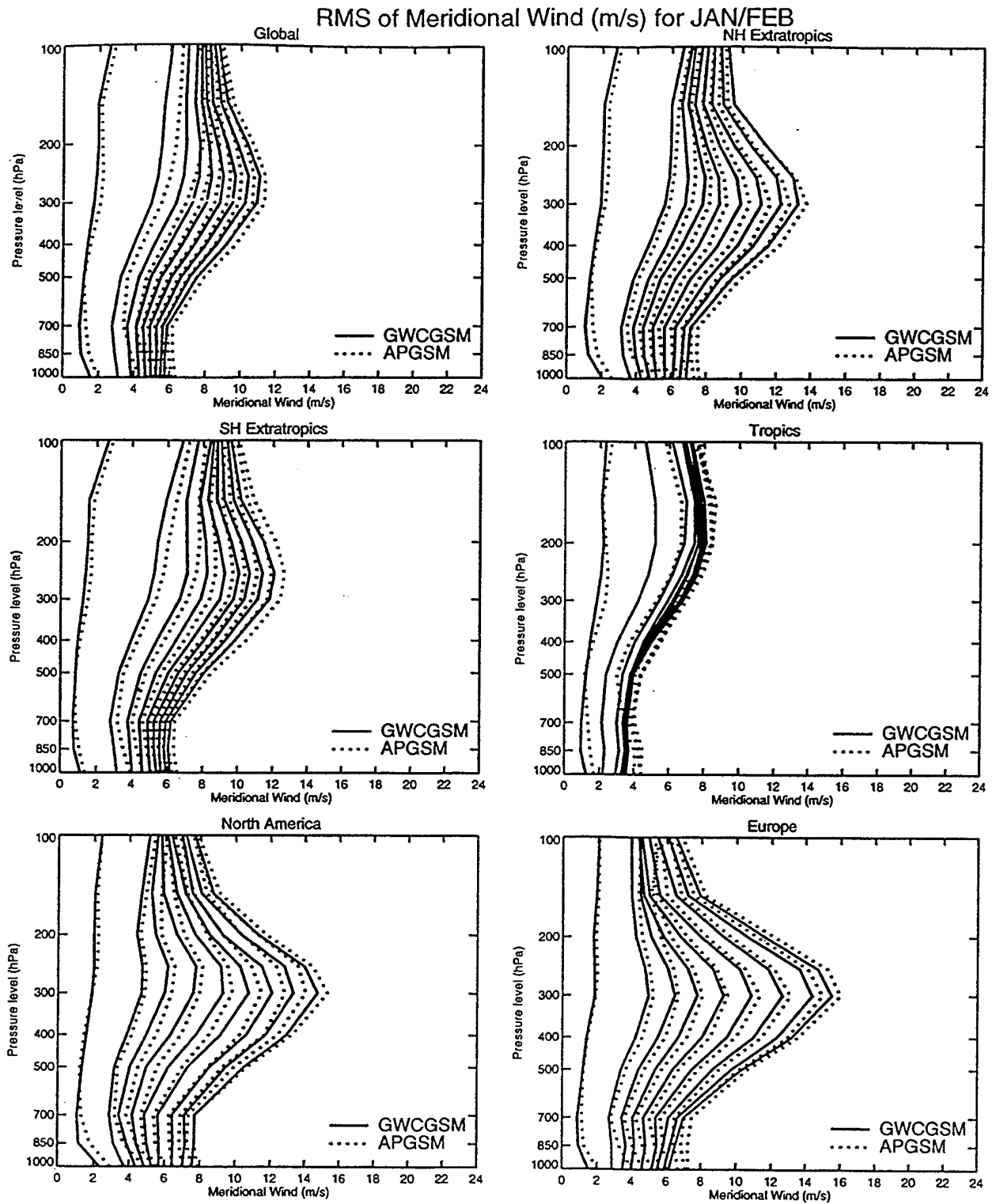


Figure 18: As Figure 5, except for the RMS errors of the meridional wind.



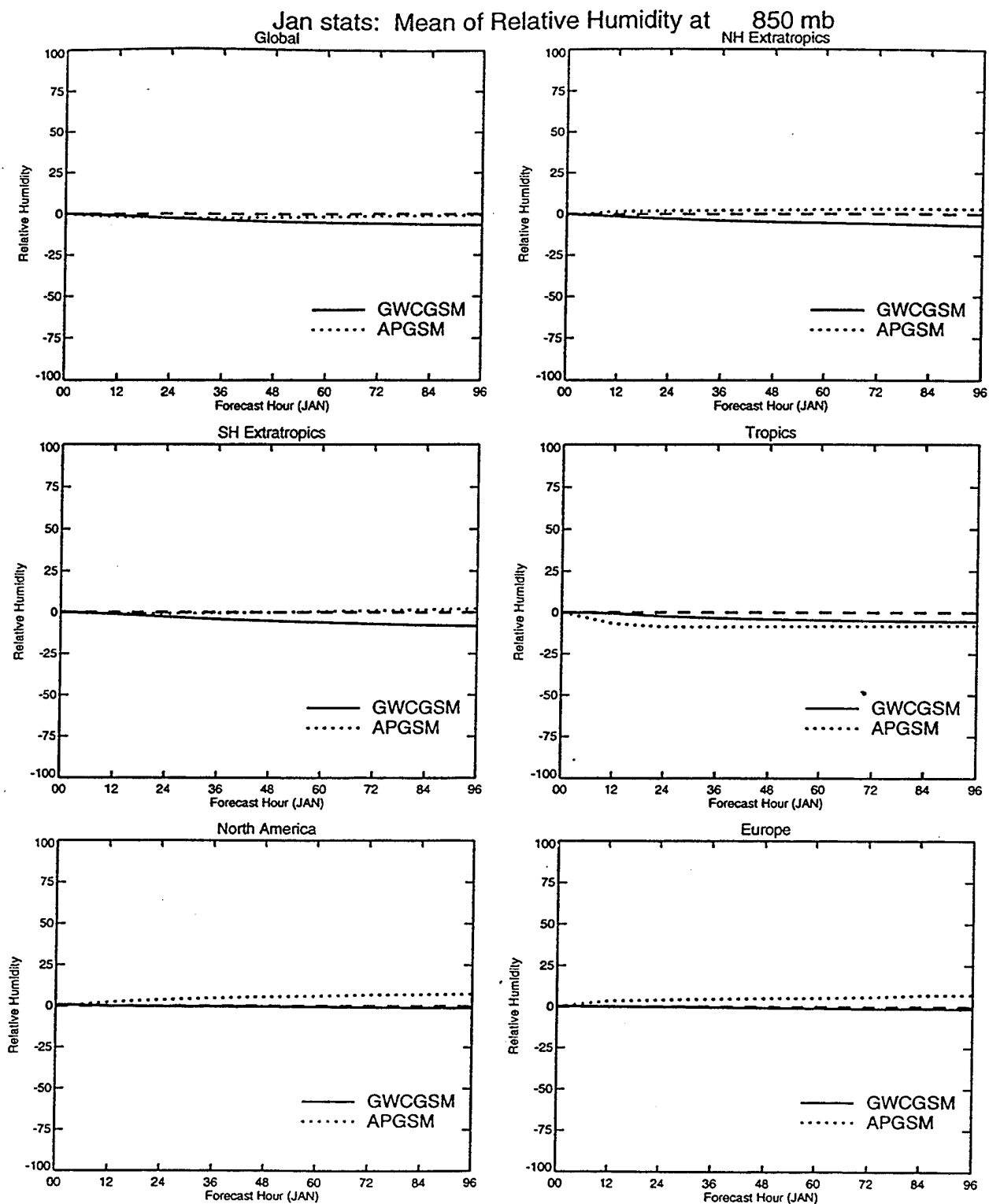


Figure 19: As Figure 2, except for the bias of 850 hPa RH.

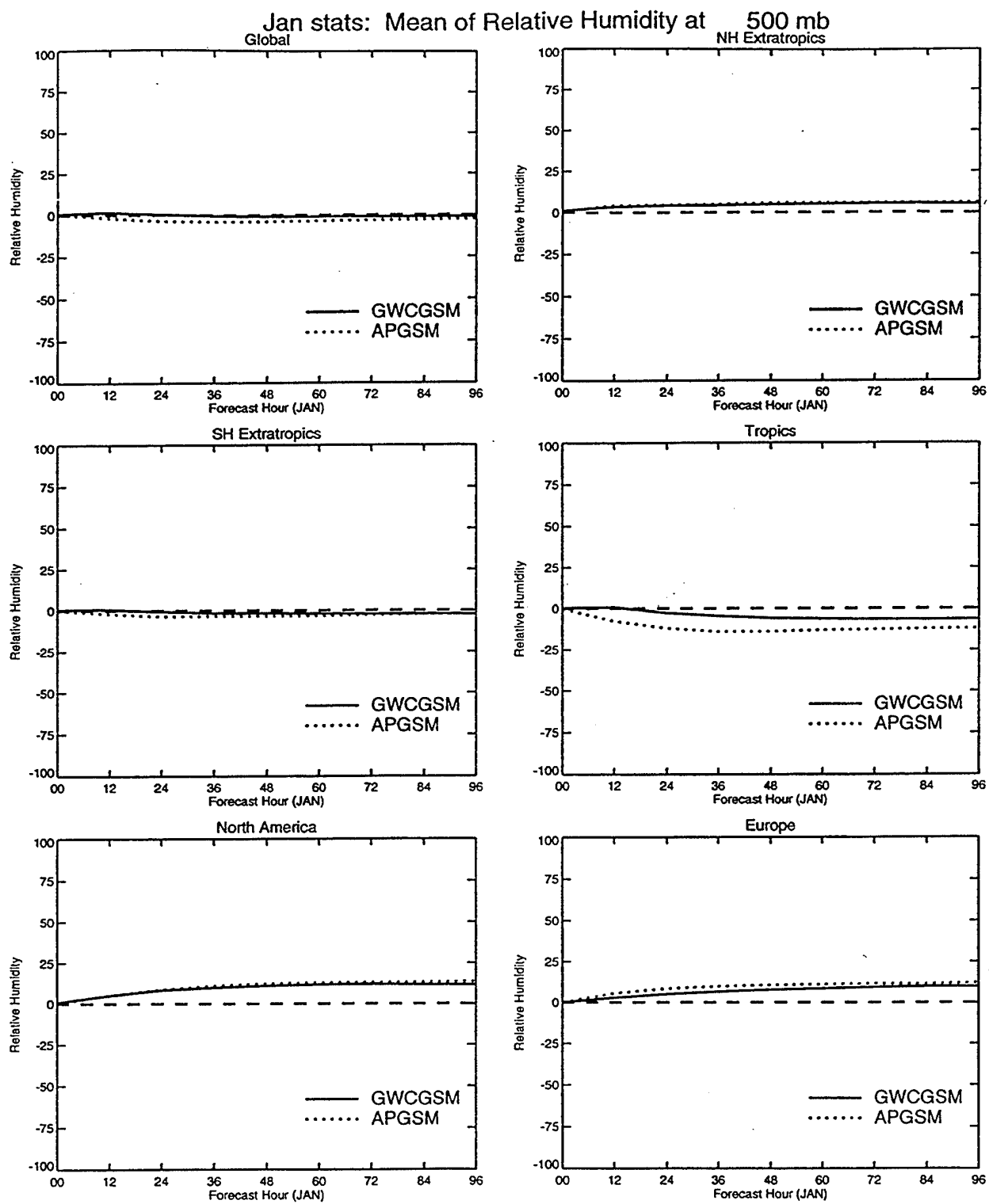


Figure 20: As Figure 2, except for the bias of 500 hPa RH.

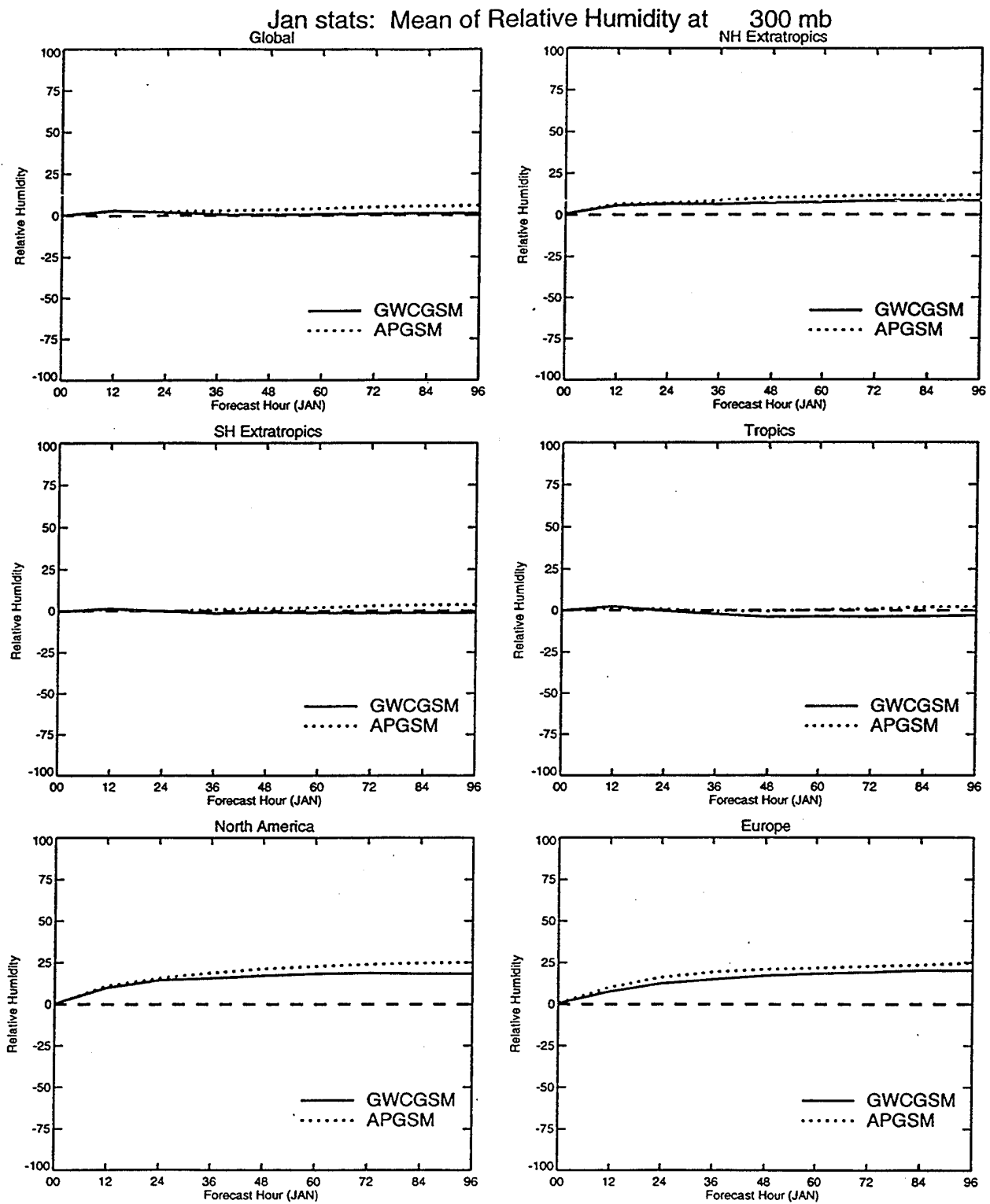


Figure 21: As Figure 2, except for the bias of 300 hPa RH.

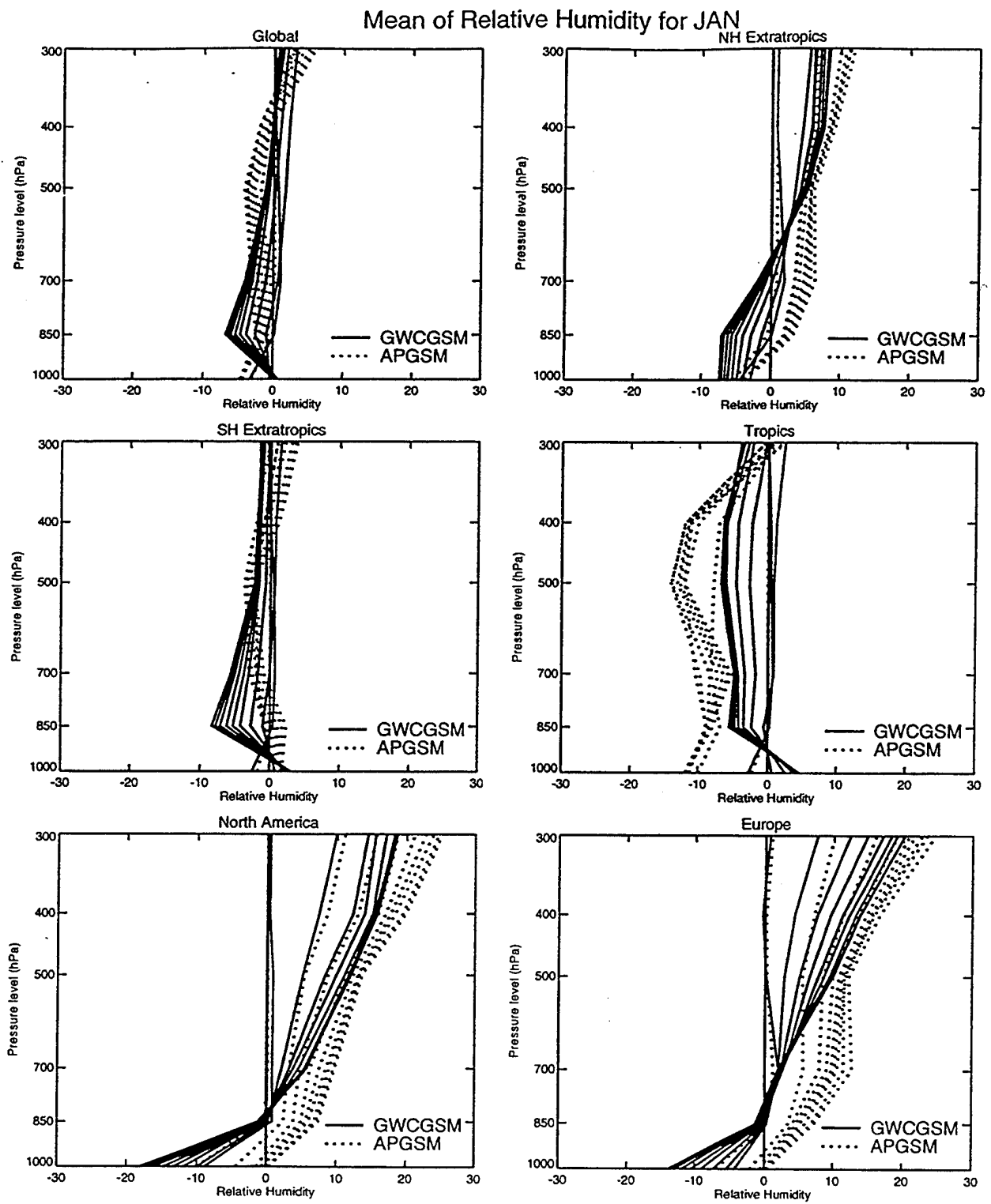


Figure 22: As Figure 5, except for the bias of RH.

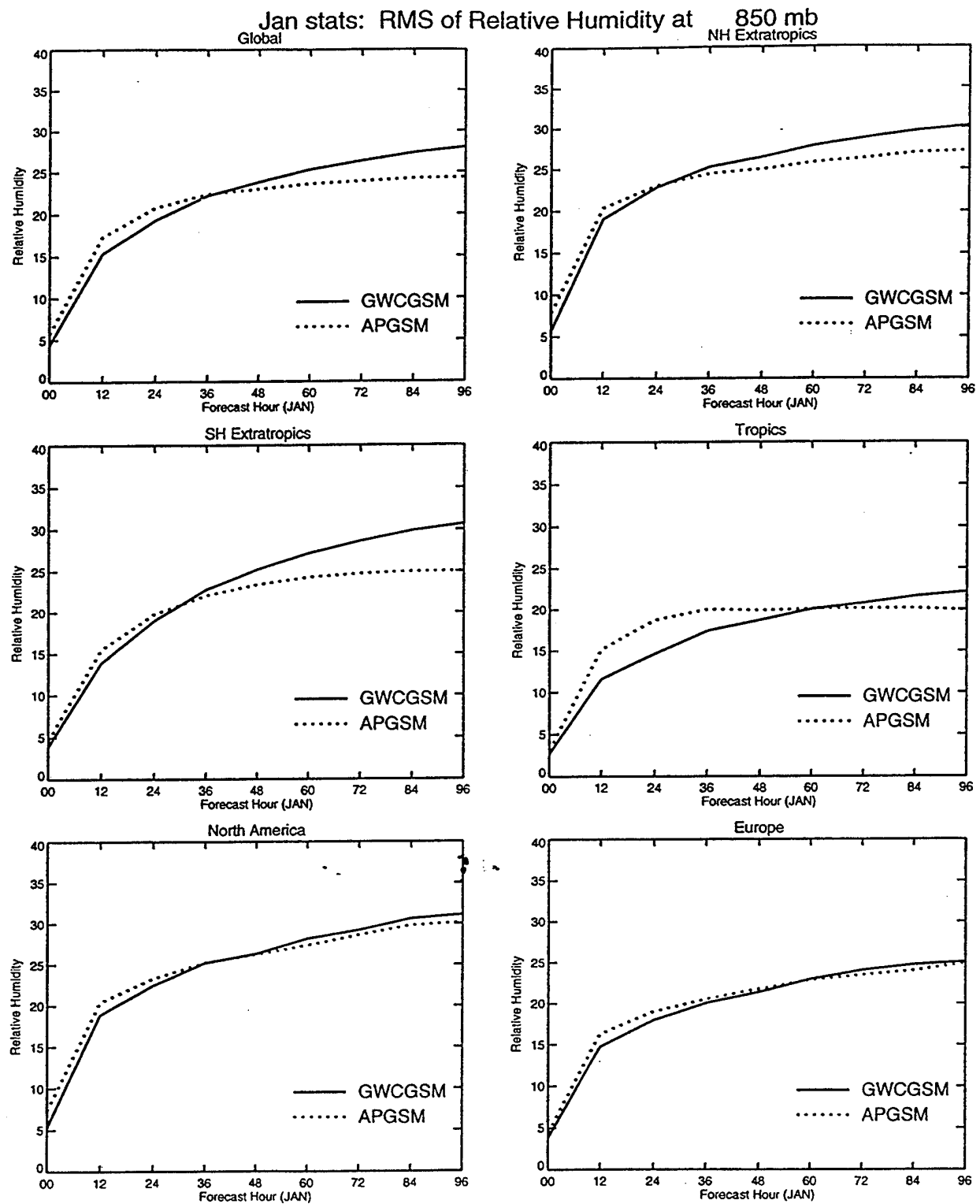


Figure 23: As Figure 2, except for the RMS error of 850 hPa RH.

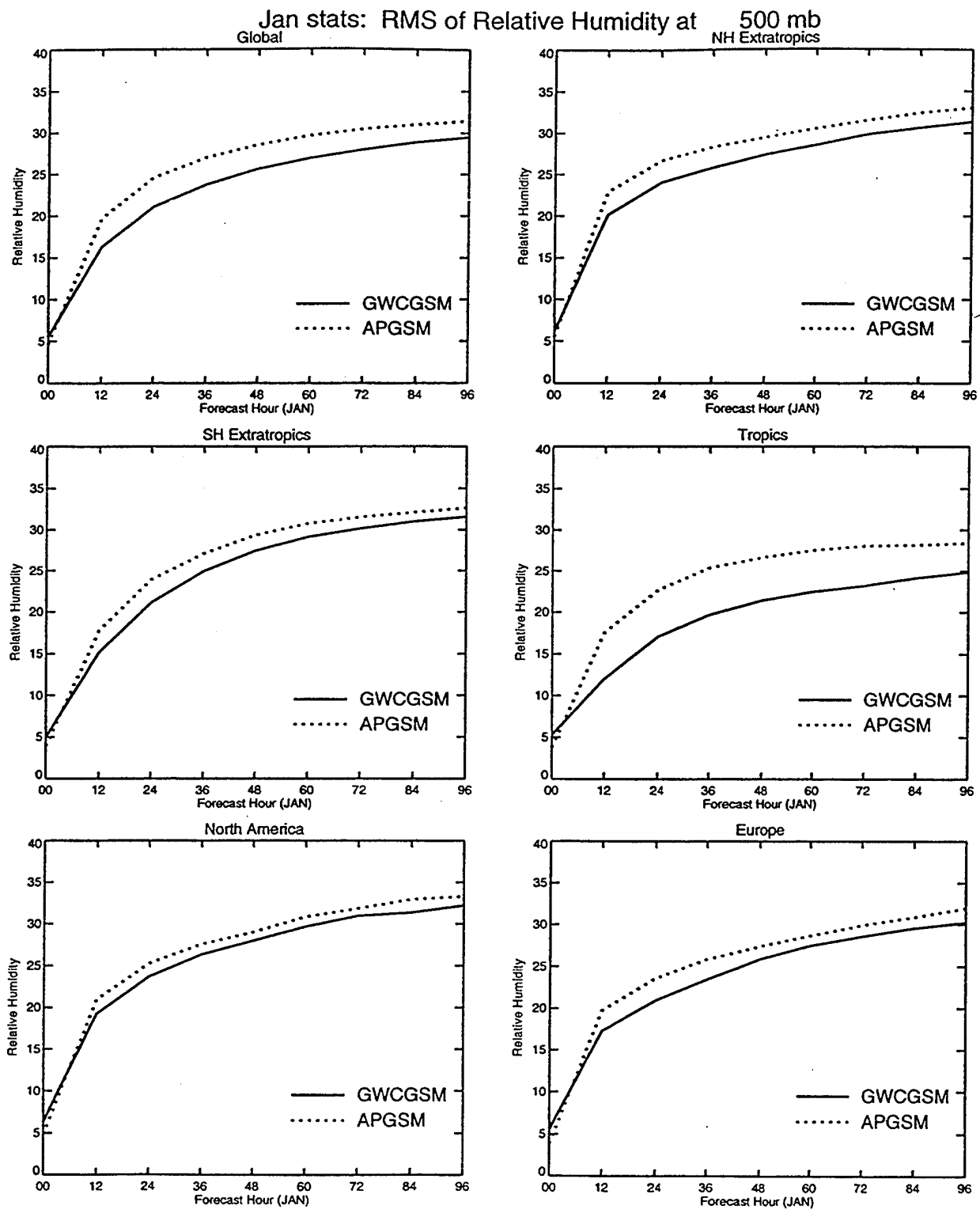


Figure 24: As Figure 2, except for the RMS error of 500 hPa RH.

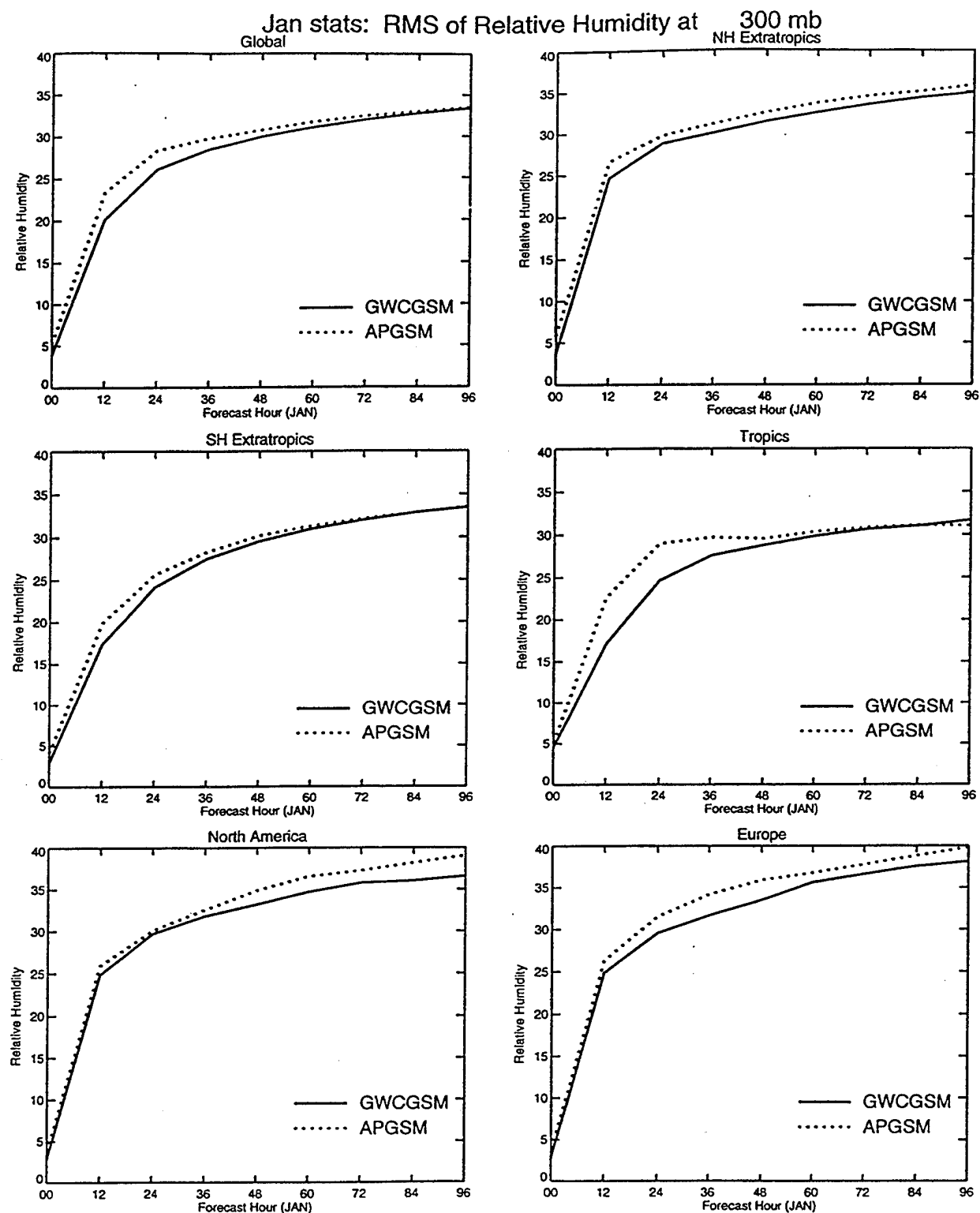


Figure 25: As Figure 2, except for the RMS error of 300 hPa RH.

# RMS of Relative Humidity for JAN

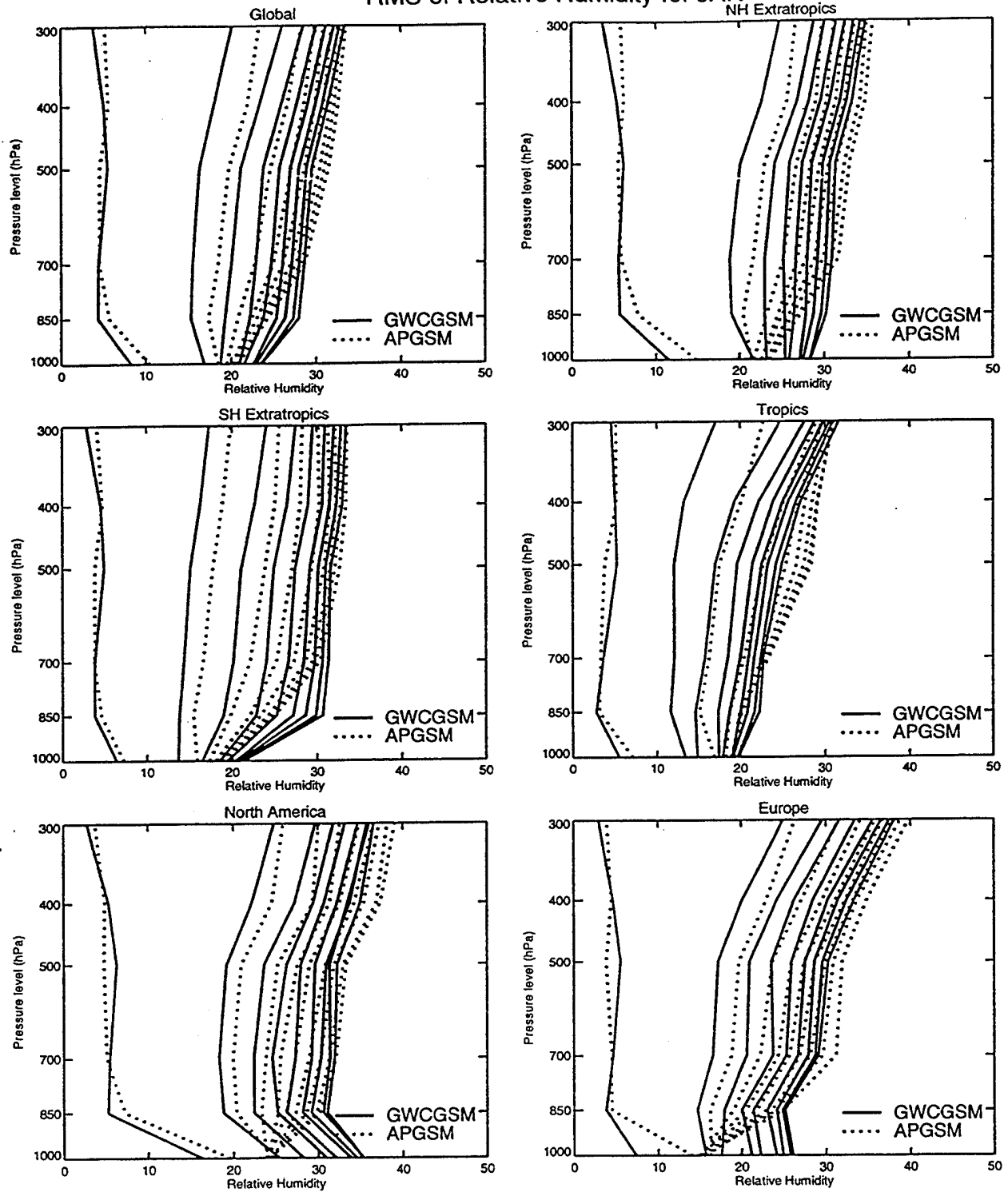


Figure 26: As Figure 5, except for the RMS error of RH.



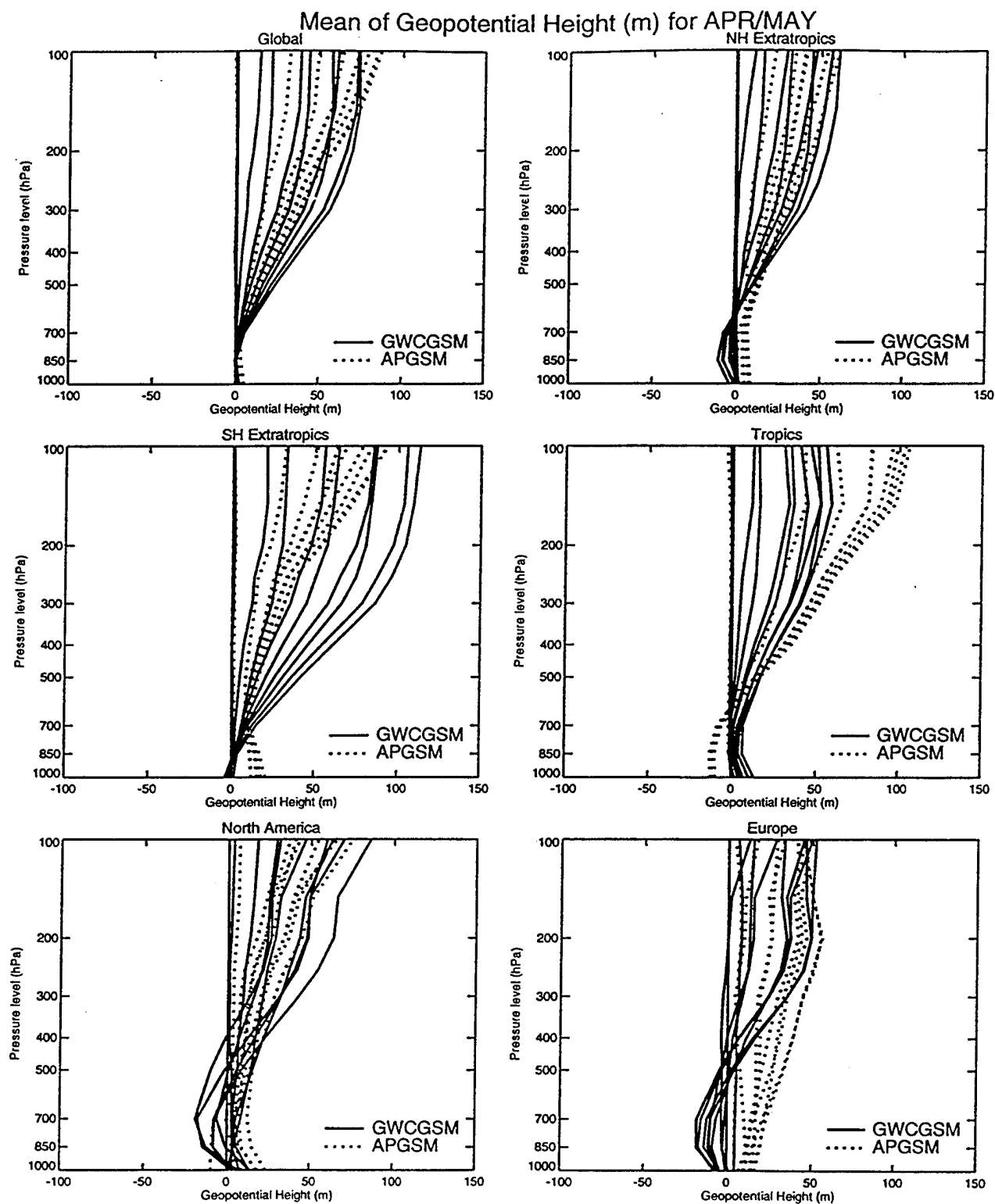


Figure 27: As Figure 5, except for the bias of geopotential for April.

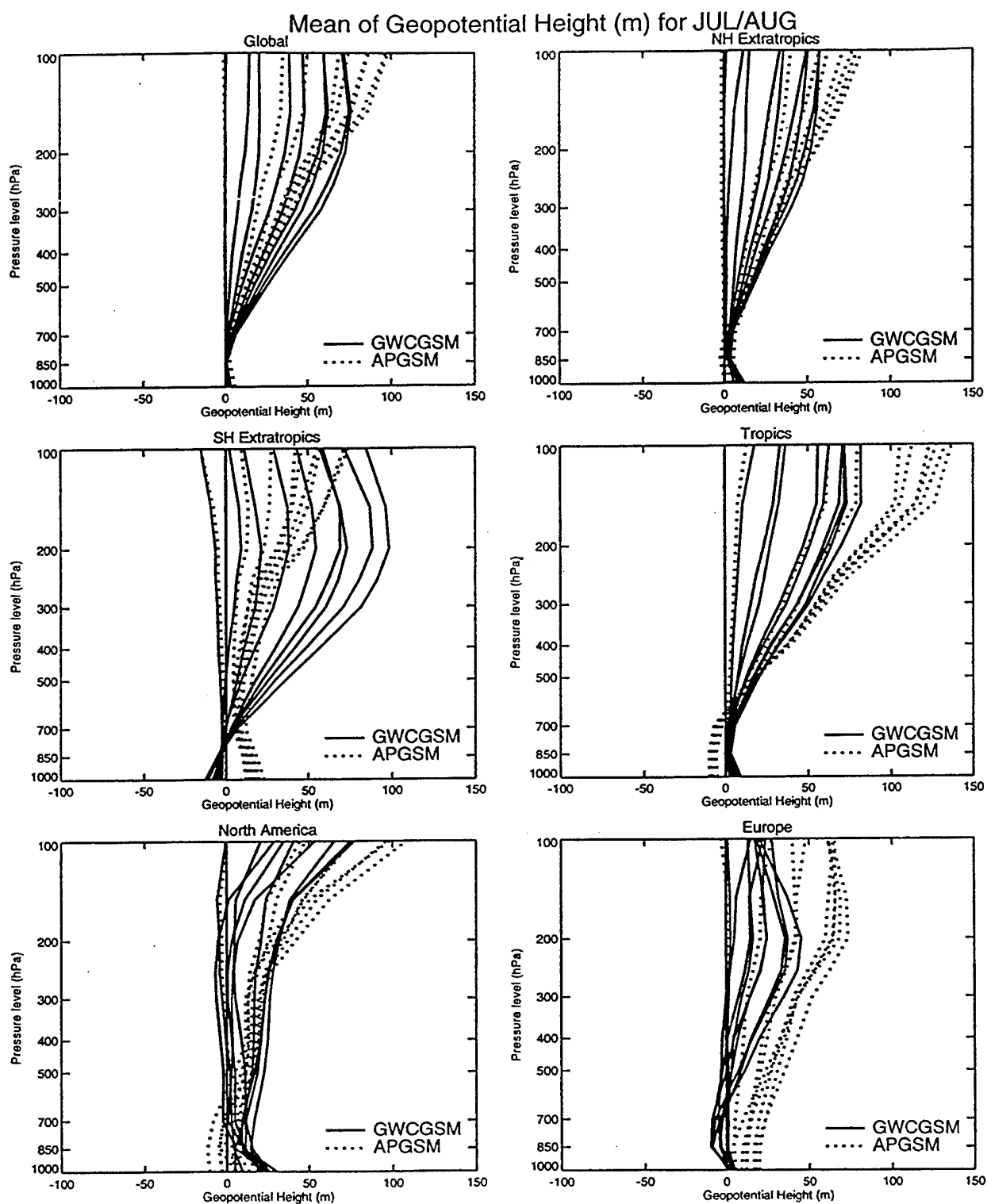


Figure 28: As Figure 5, except for the bias of geopotential for July.

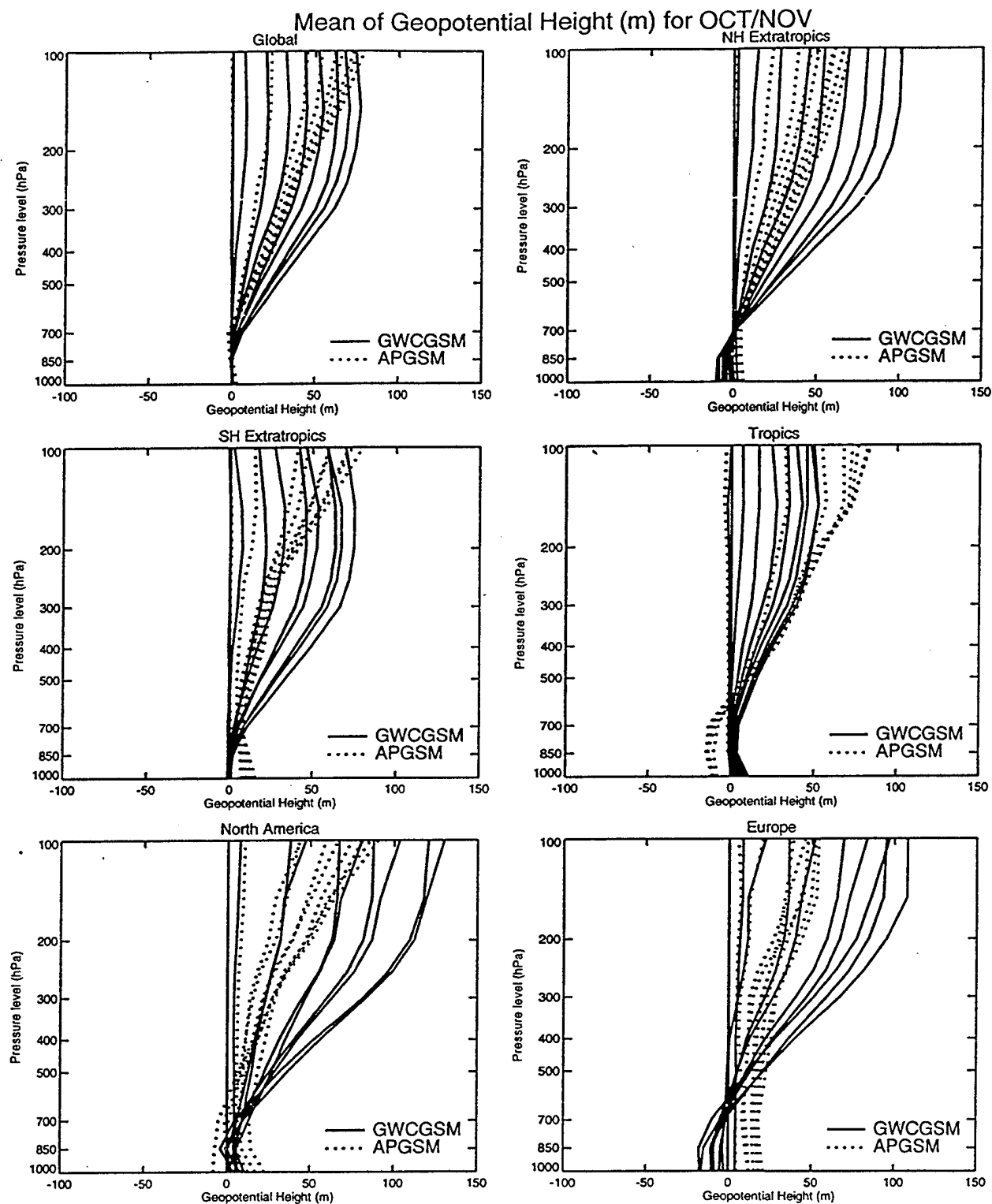


Figure 29: As Figure 5, except for the bias of geopotential for October.

# Apr stats: RMS of Geopotential Height (m) at 1000 mb

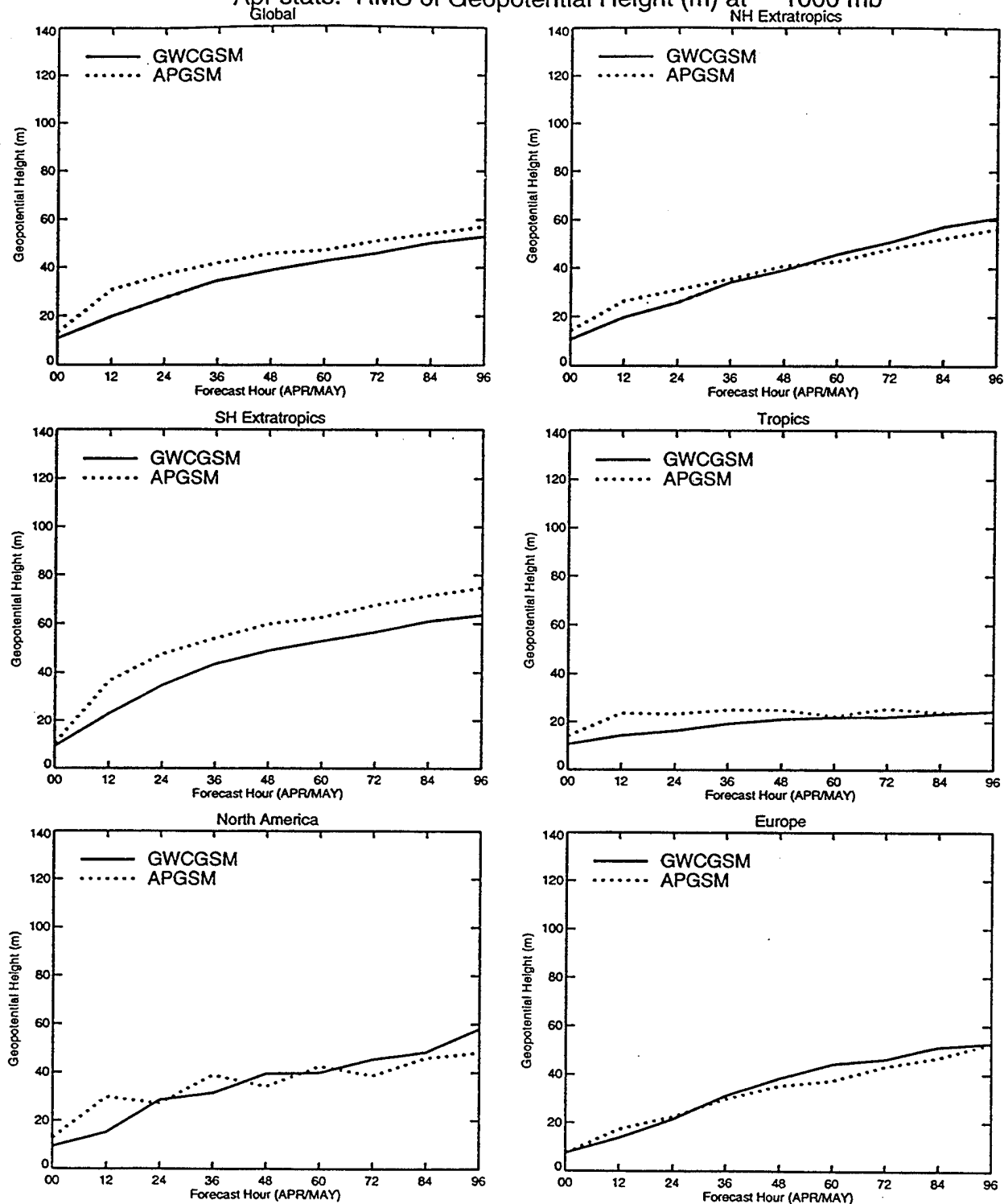


Figure 30: As Figure 2, except for the RMS error of 1000 hPa geopotential height for April.

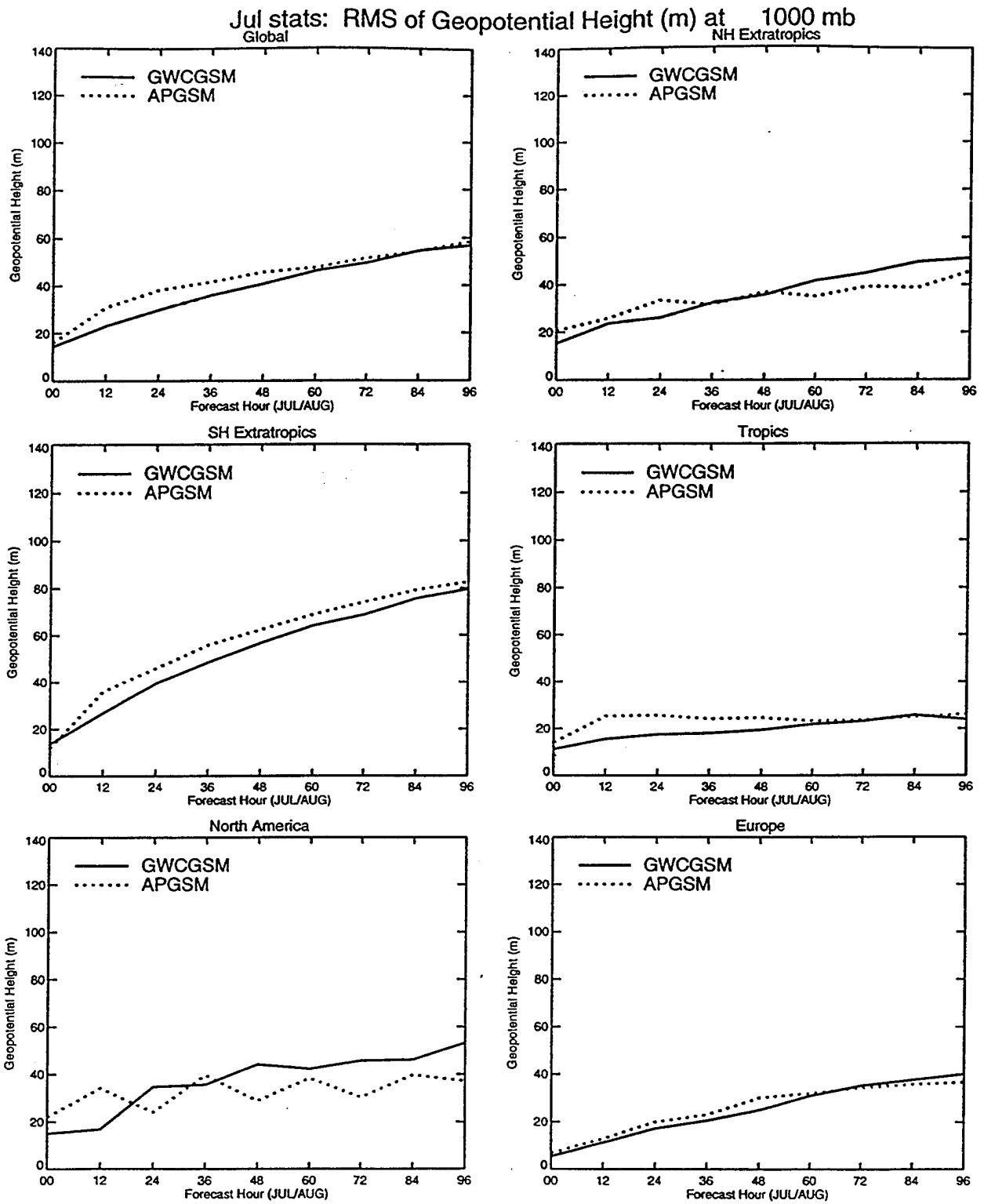


Figure 31: As Figure 2, except for the RMS error of 1000 hPa geopotential height for July.

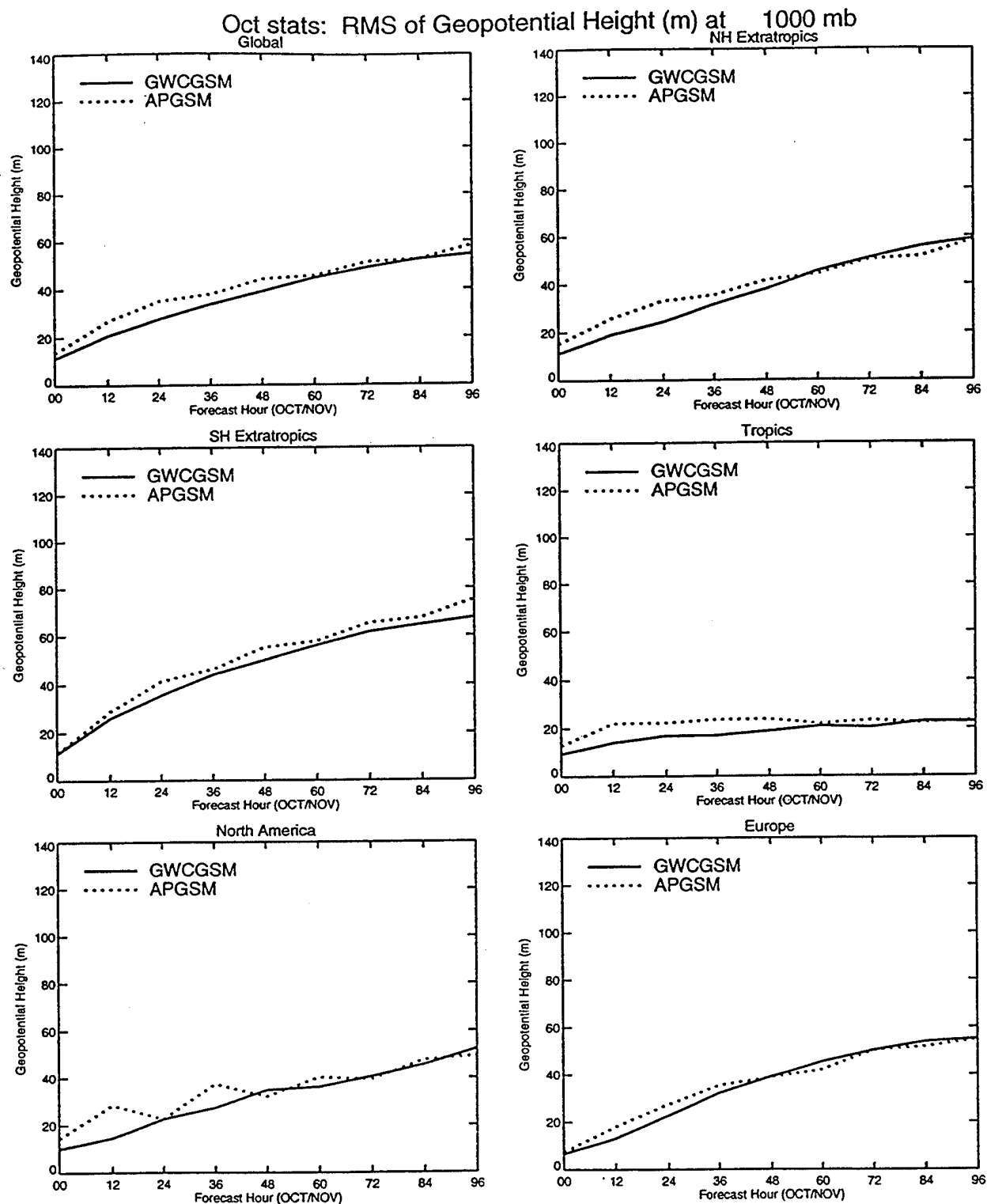


Figure 32: As Figure 2, except for the RMS error of 1000 hPa geopotential height for October.

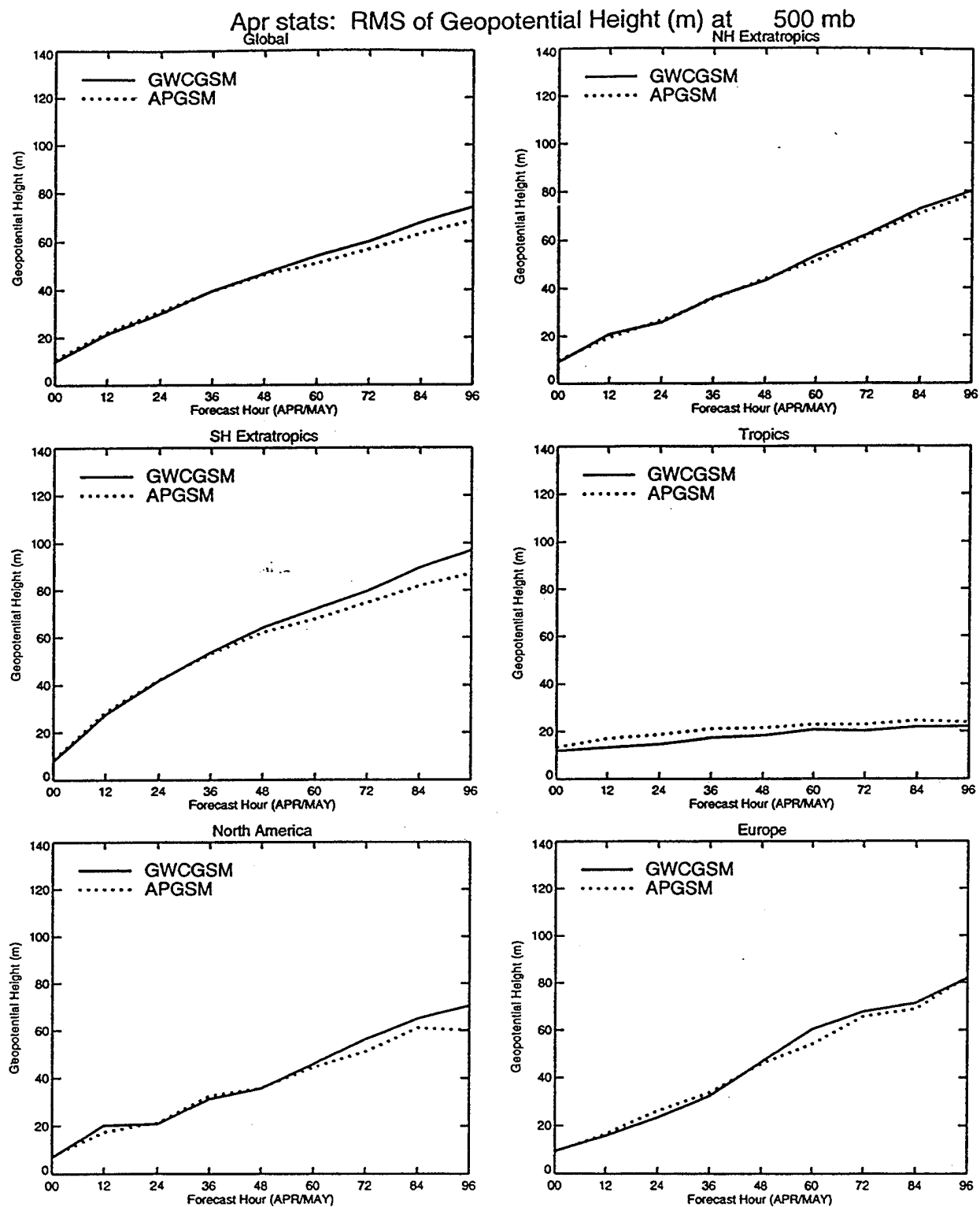


Figure 33: As Figure 2, except for the RMS error of 500 hPa geopotential height for April.

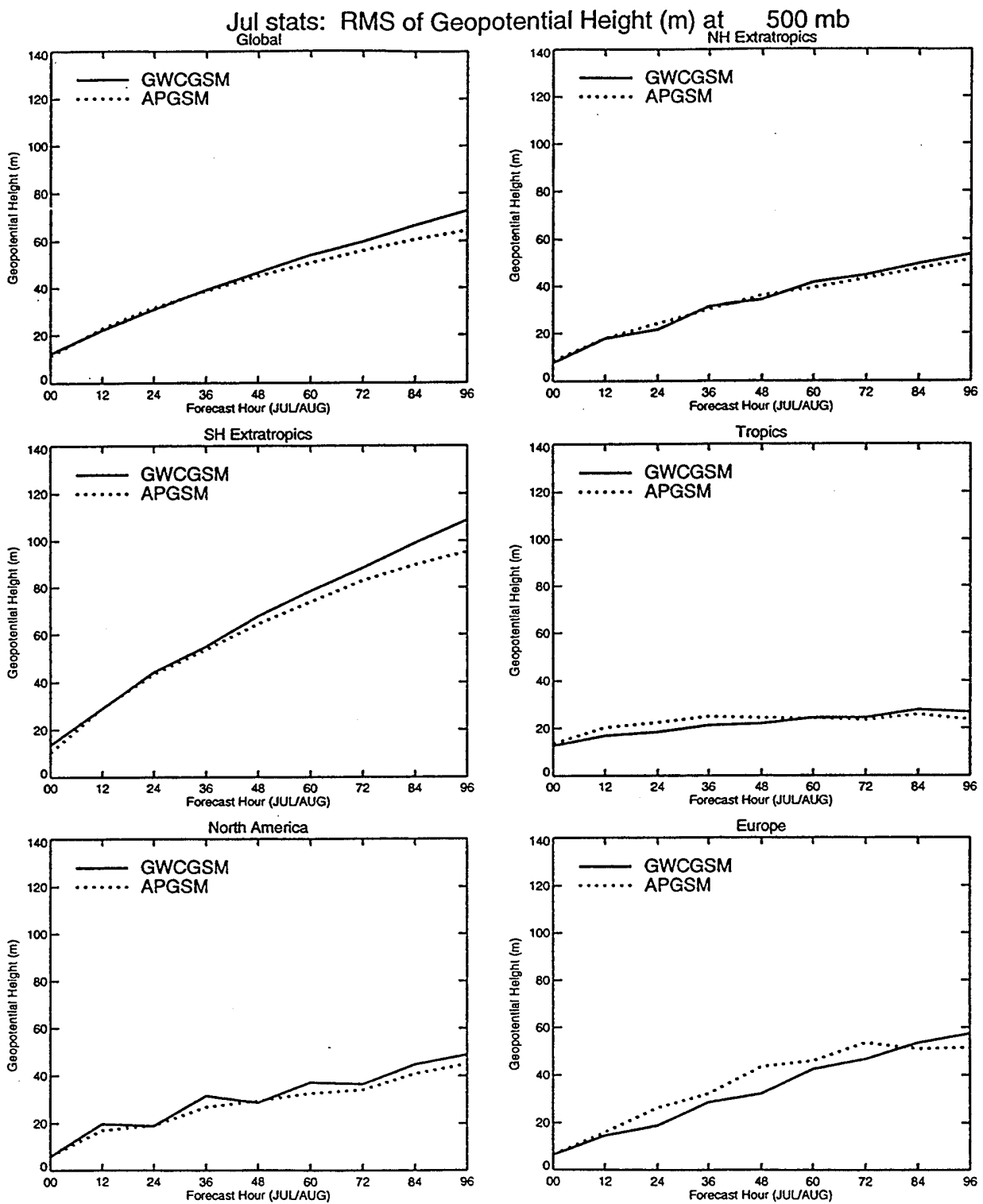


Figure 34: As Figure 2, except for the RMS error of 500 hPa geopotential height for July.



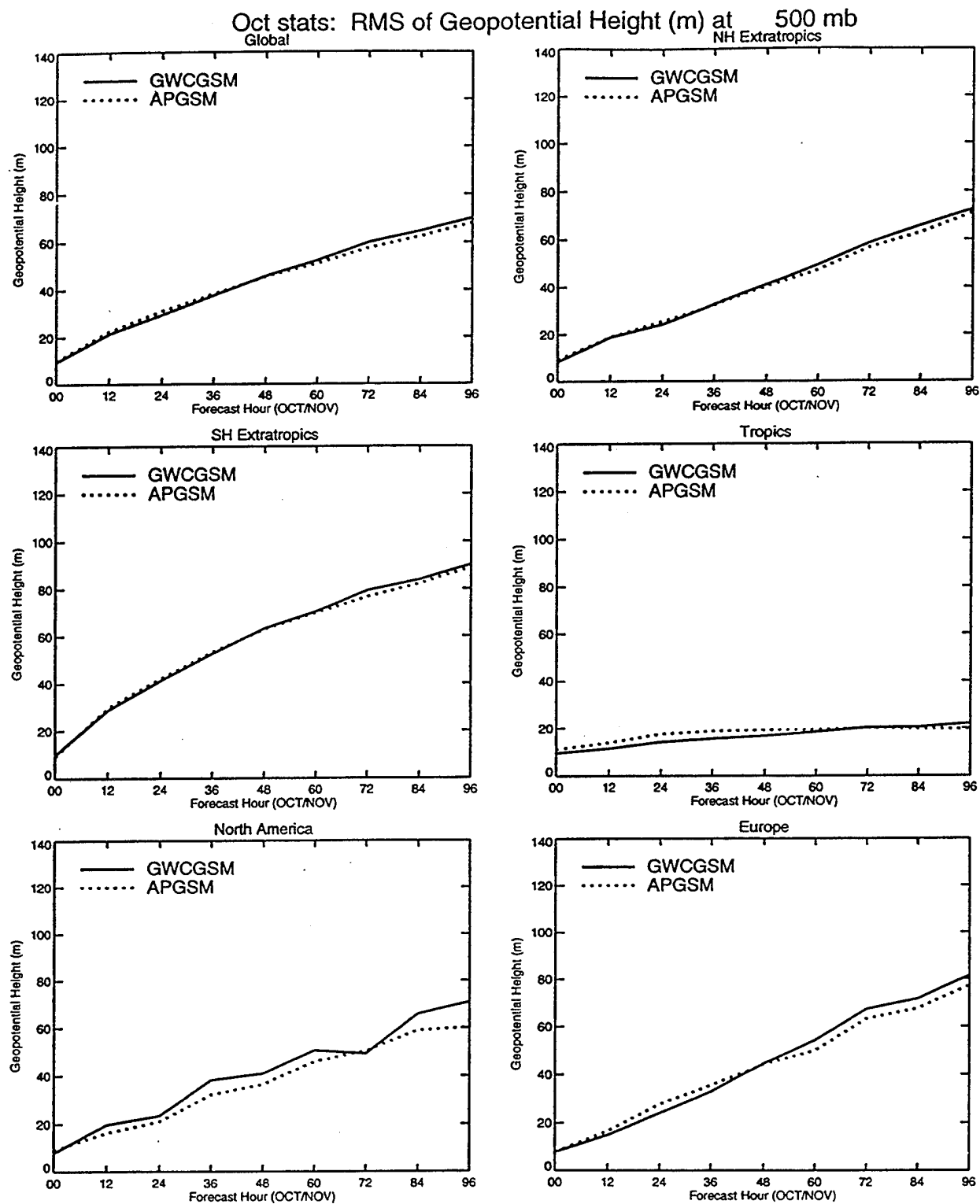


Figure 35: As Figure 2, except for the RMS error of 500 hPa geopotential height for October.

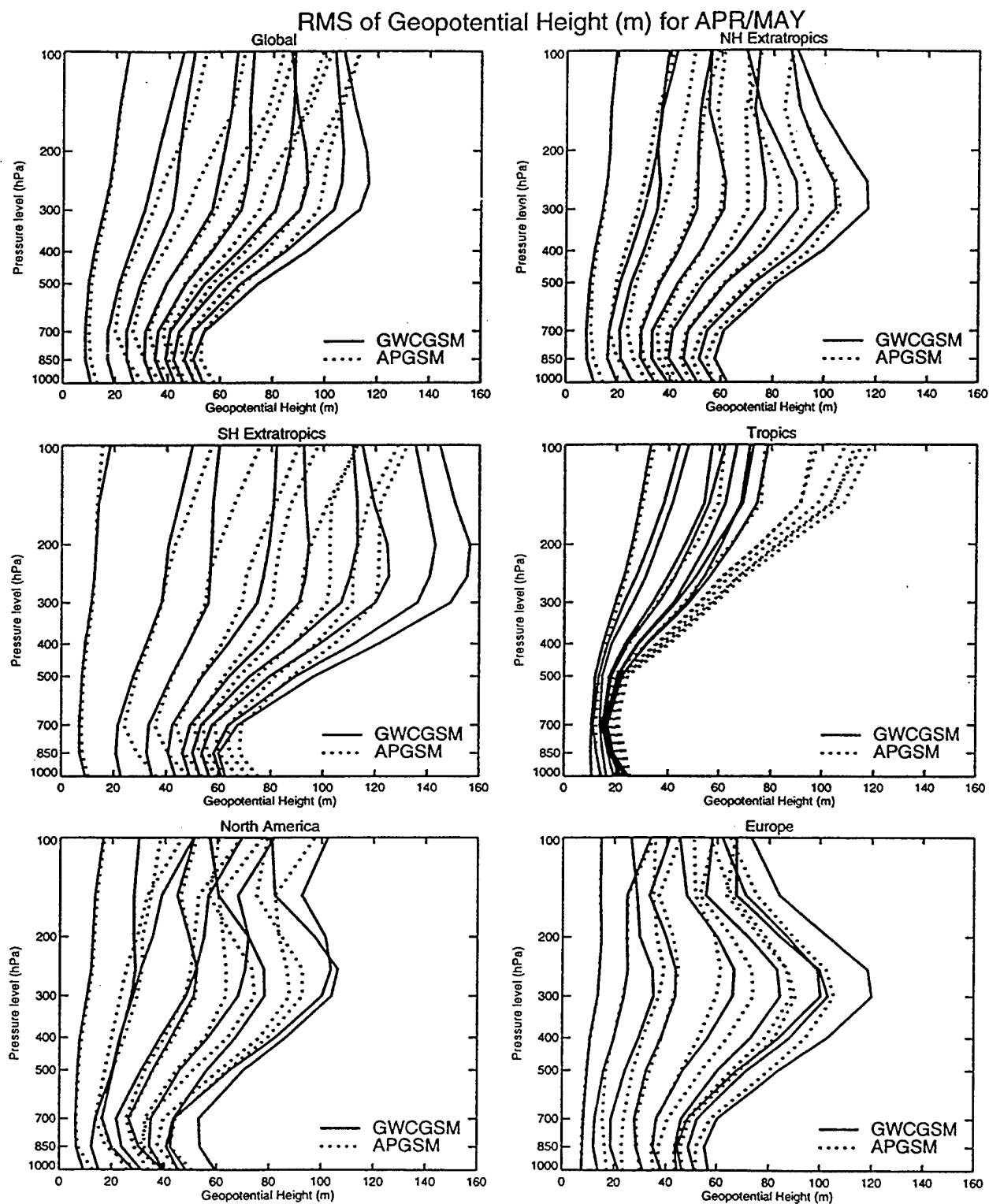


Figure 36: As Figure 5, except for the RMS errors of geopotential for April.

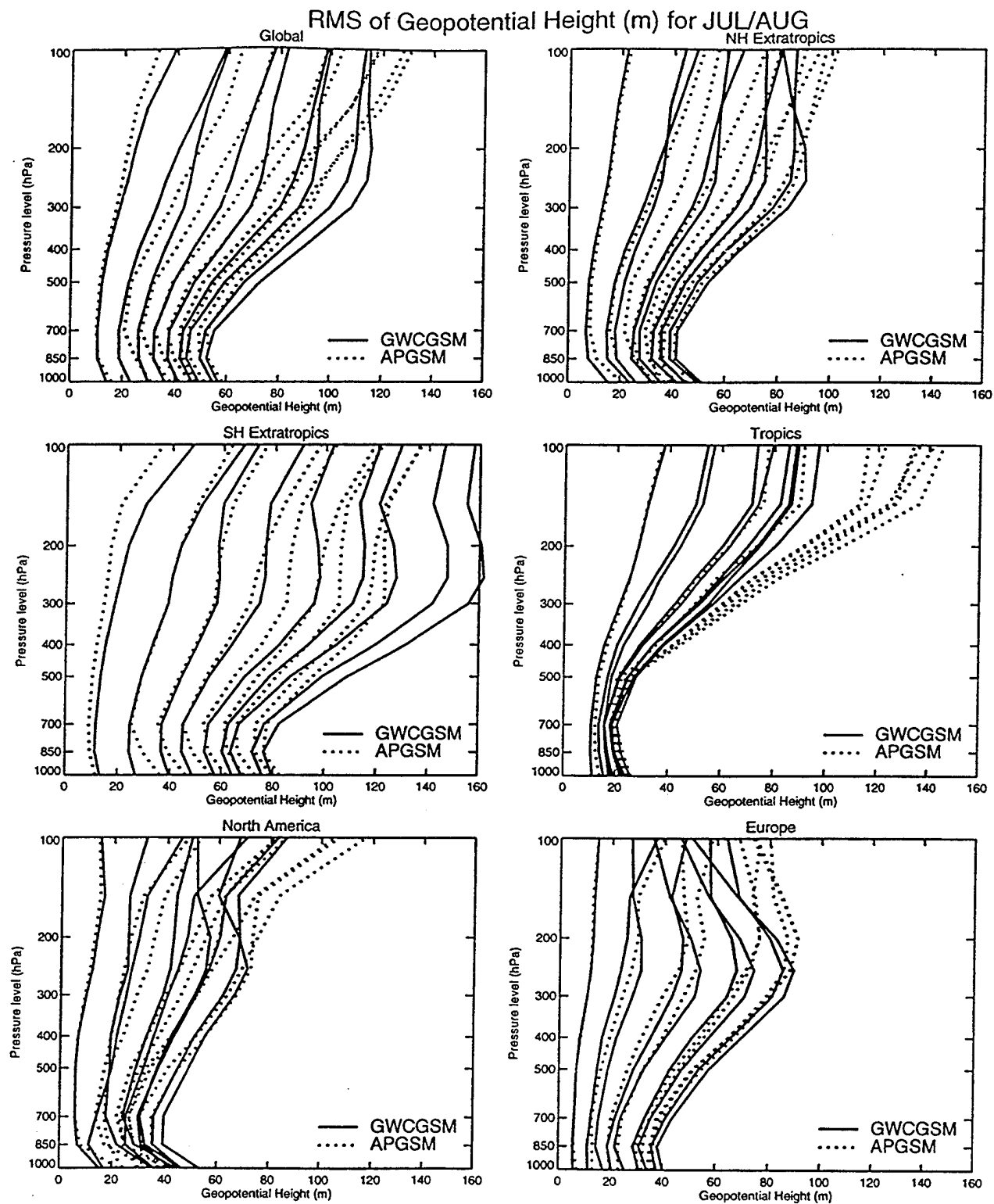


Figure 37: As Figure 5, except for the RMS errors of geopotential for July.

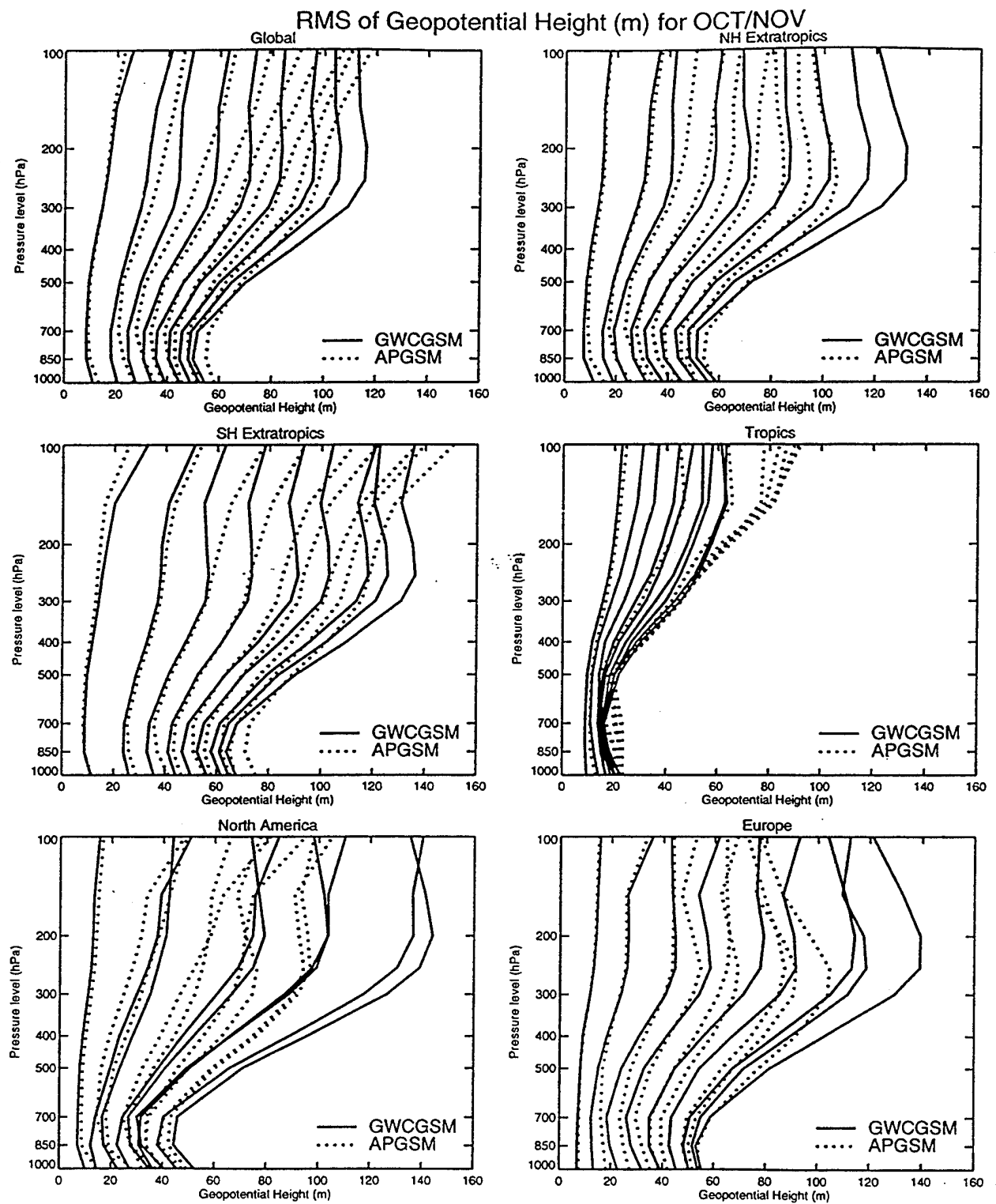


Figure 38: As Figure 5, except for the RMS errors of geopotential for October.

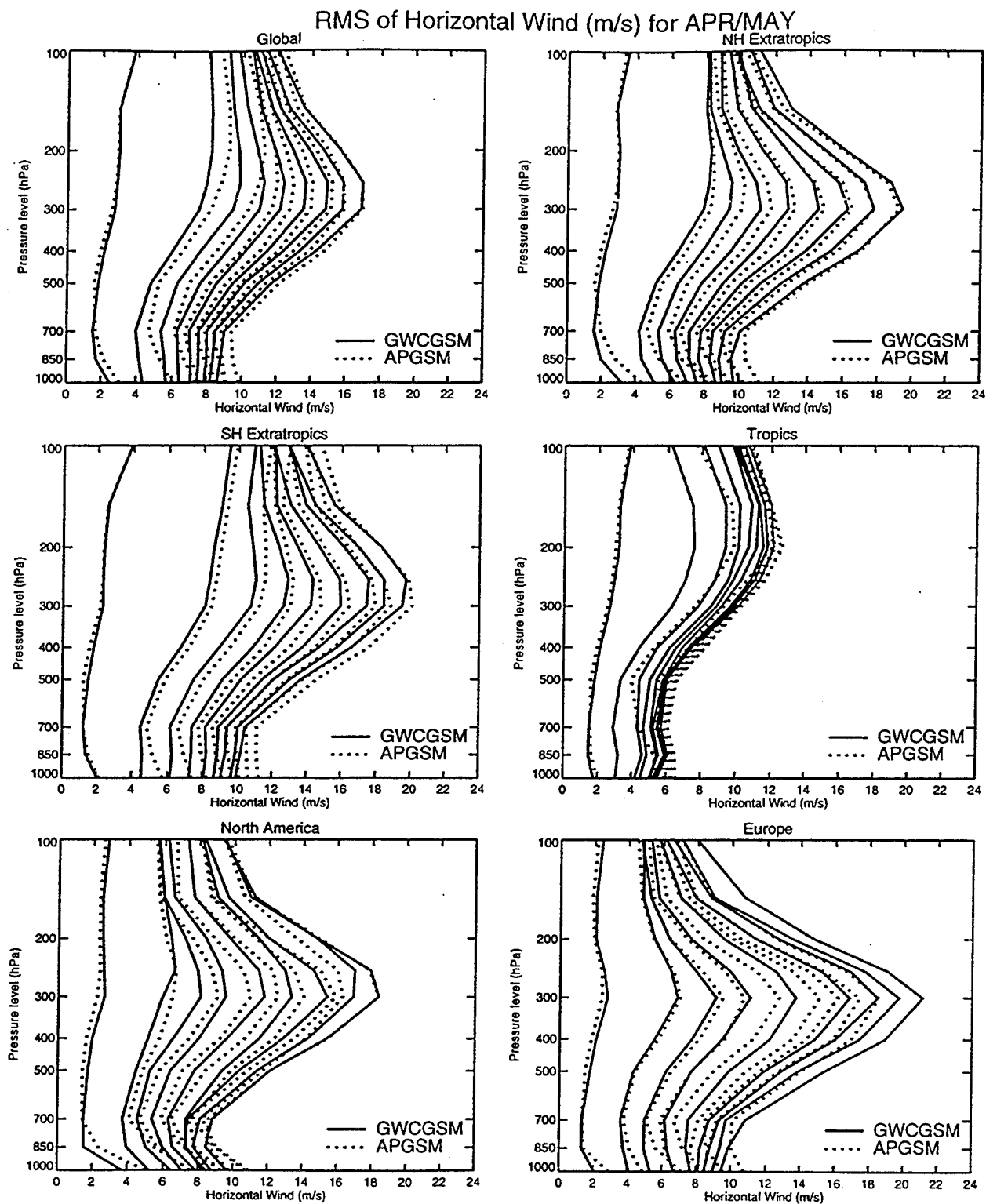


Figure 39: As Figure 5, except for the RMS errors of horizontal wind for April.

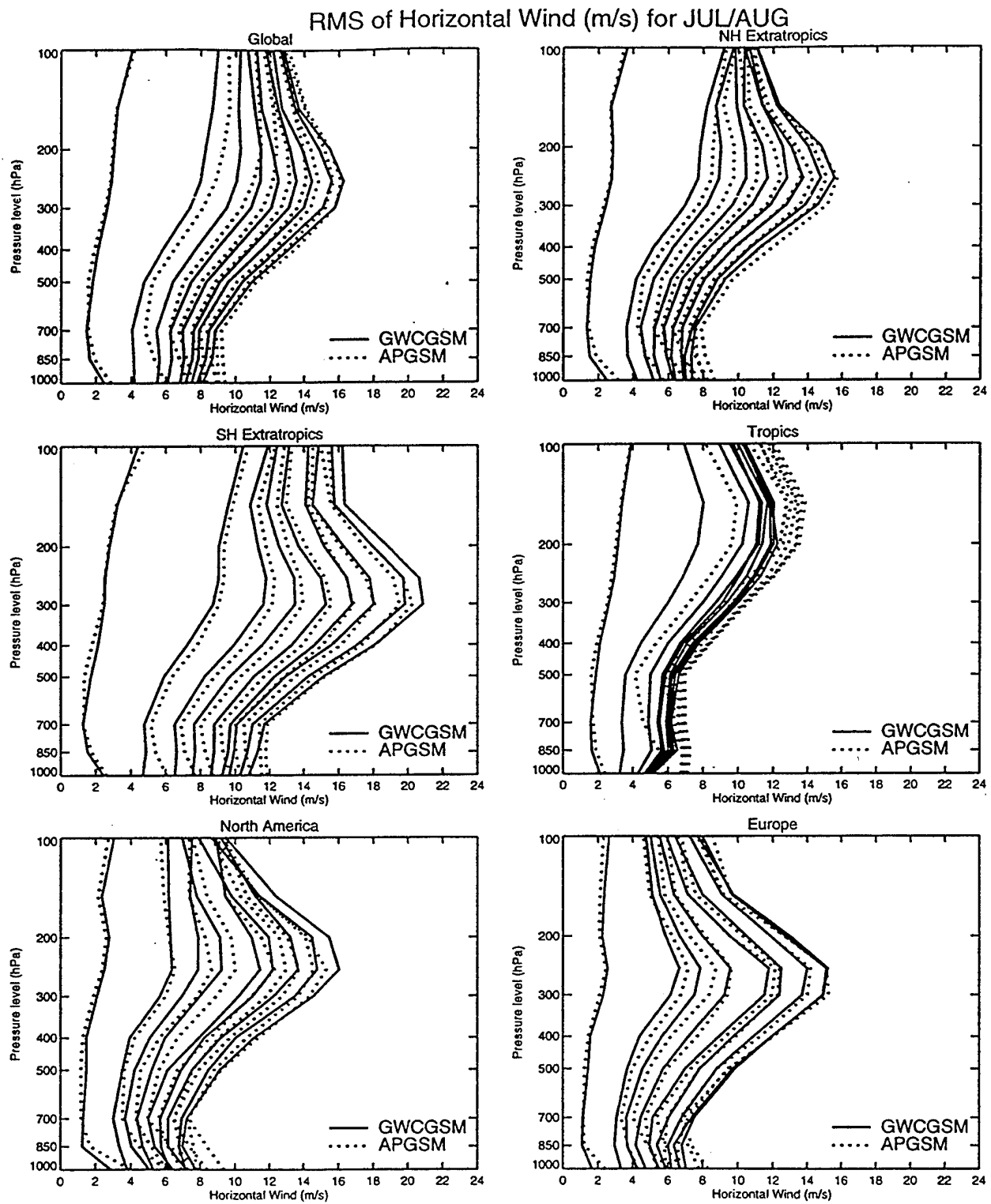


Figure 40: As Figure 5, except for the RMS errors of horizontal wind for July.

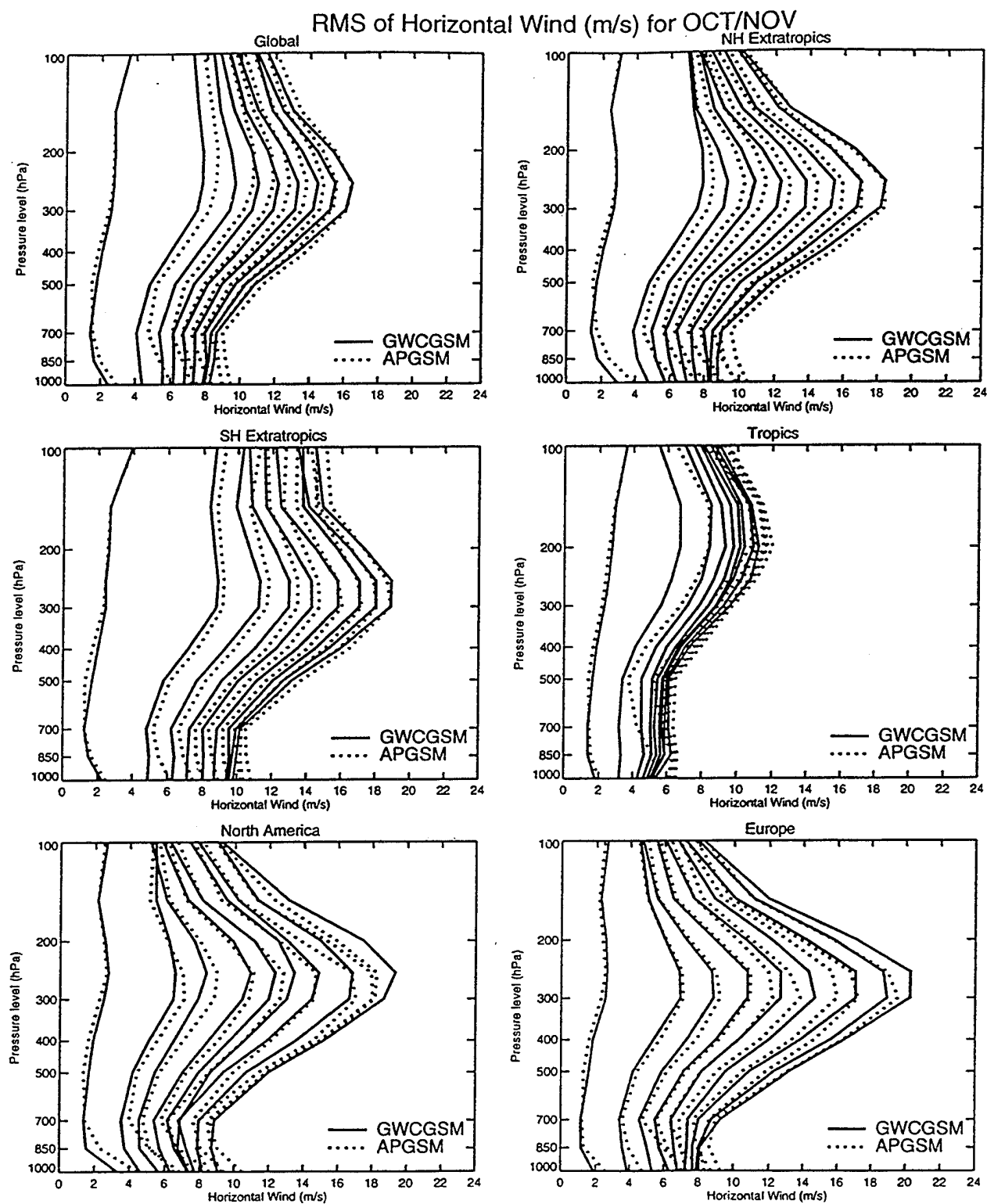


Figure 41: As Figure 5, except for the RMS errors of horizontal wind for October.

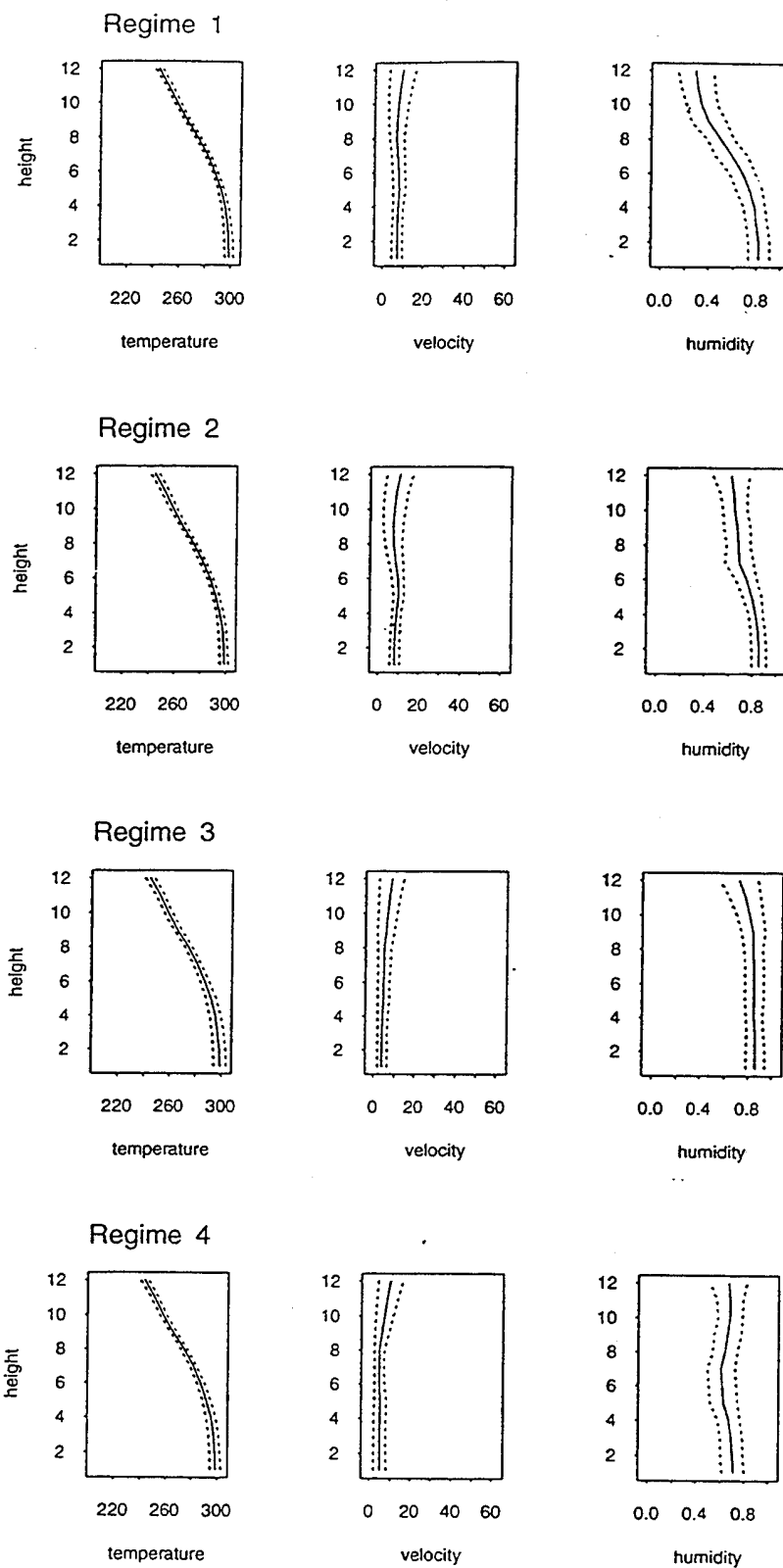


Figure 42: Mean profiles of the 24 original synoptic regimes. Shown are the temperature, wind speed, and RH, as a function of model sigma layer (going from 1 near the surface to 12 at  $\sigma=0.325$  (viz. Figure 1)).



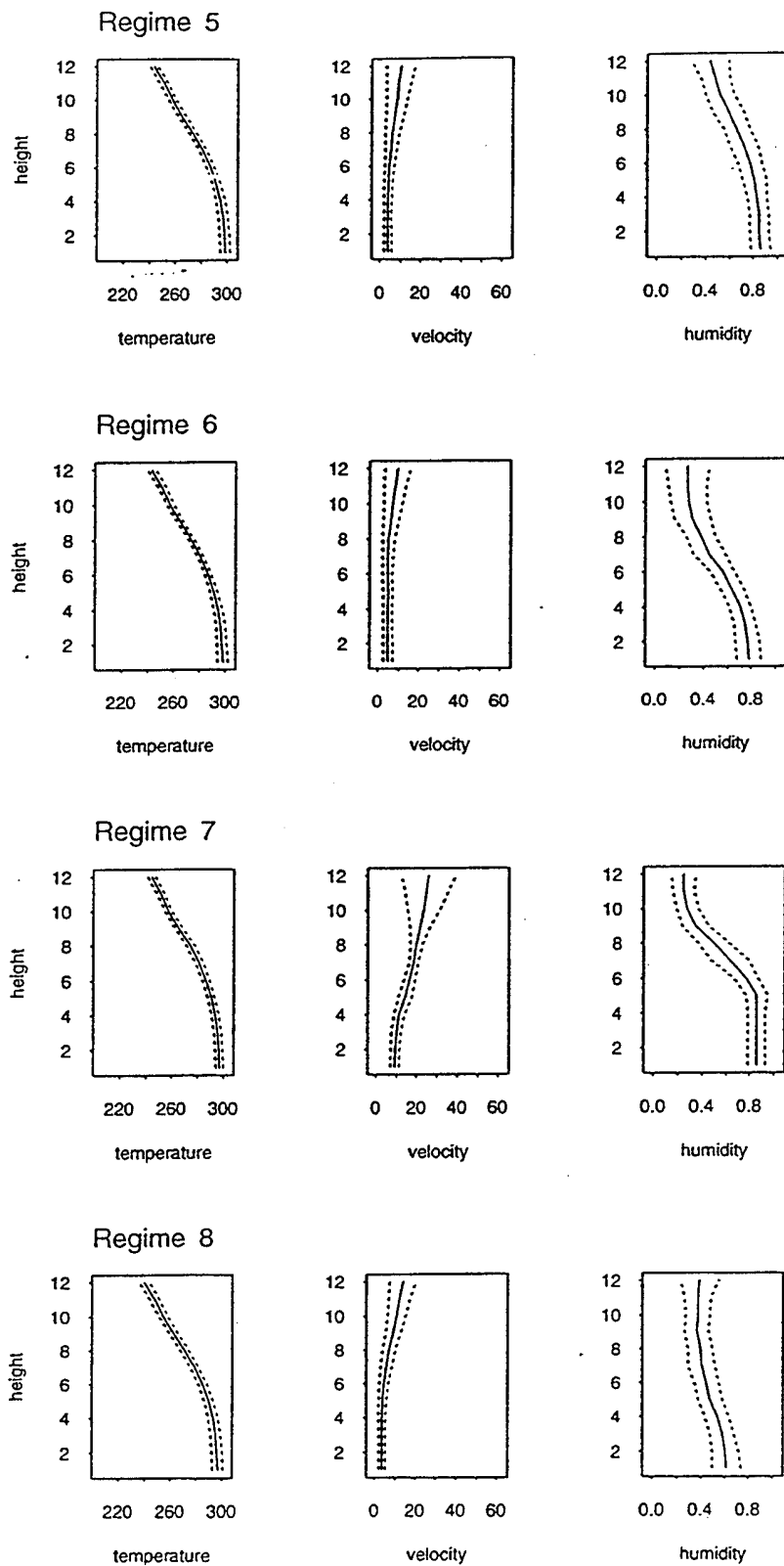


Figure 42: Continued.

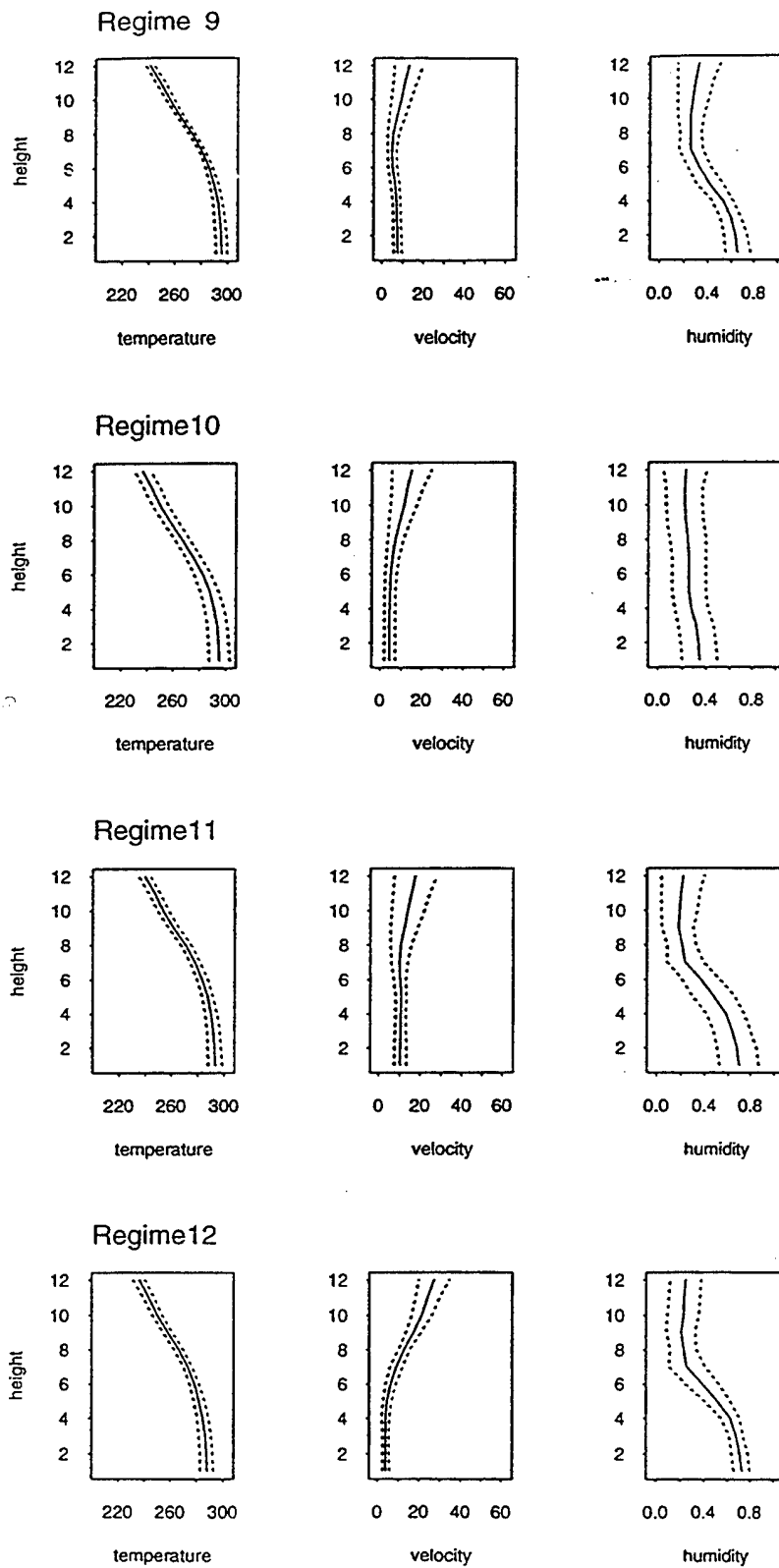


Figure 42: Continued.

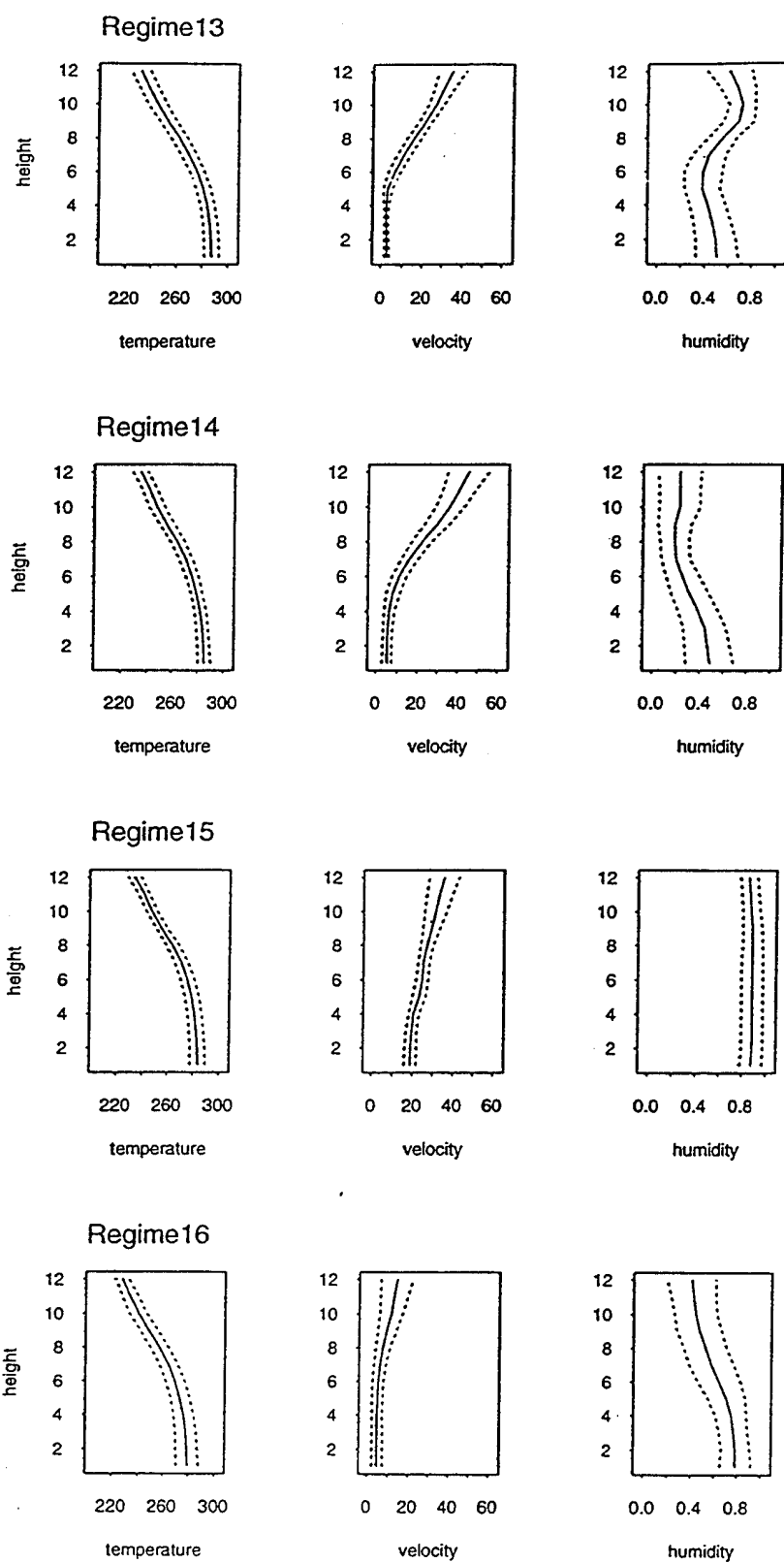


Figure 42: Continued.

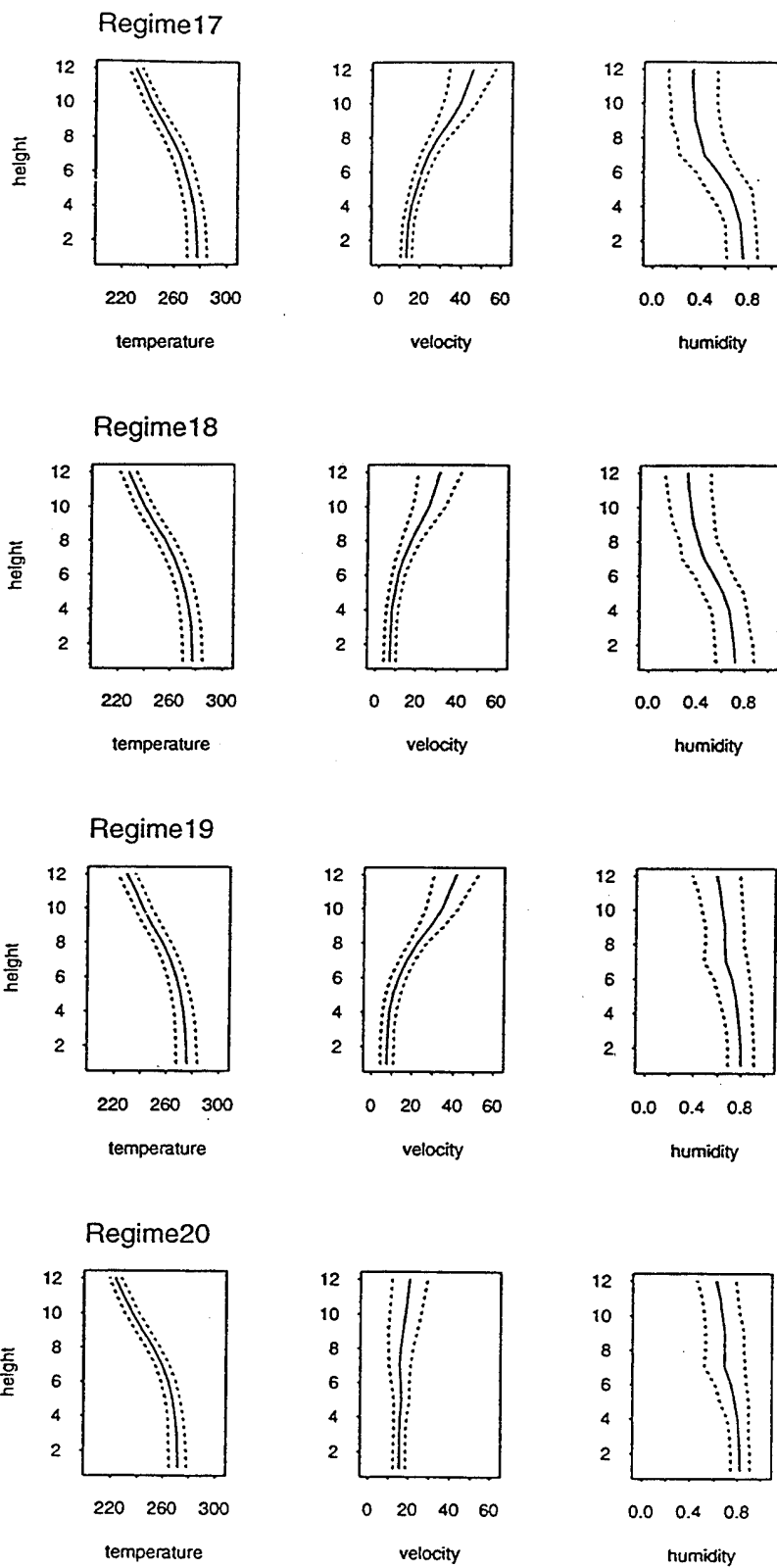


Figure 42: Continued.

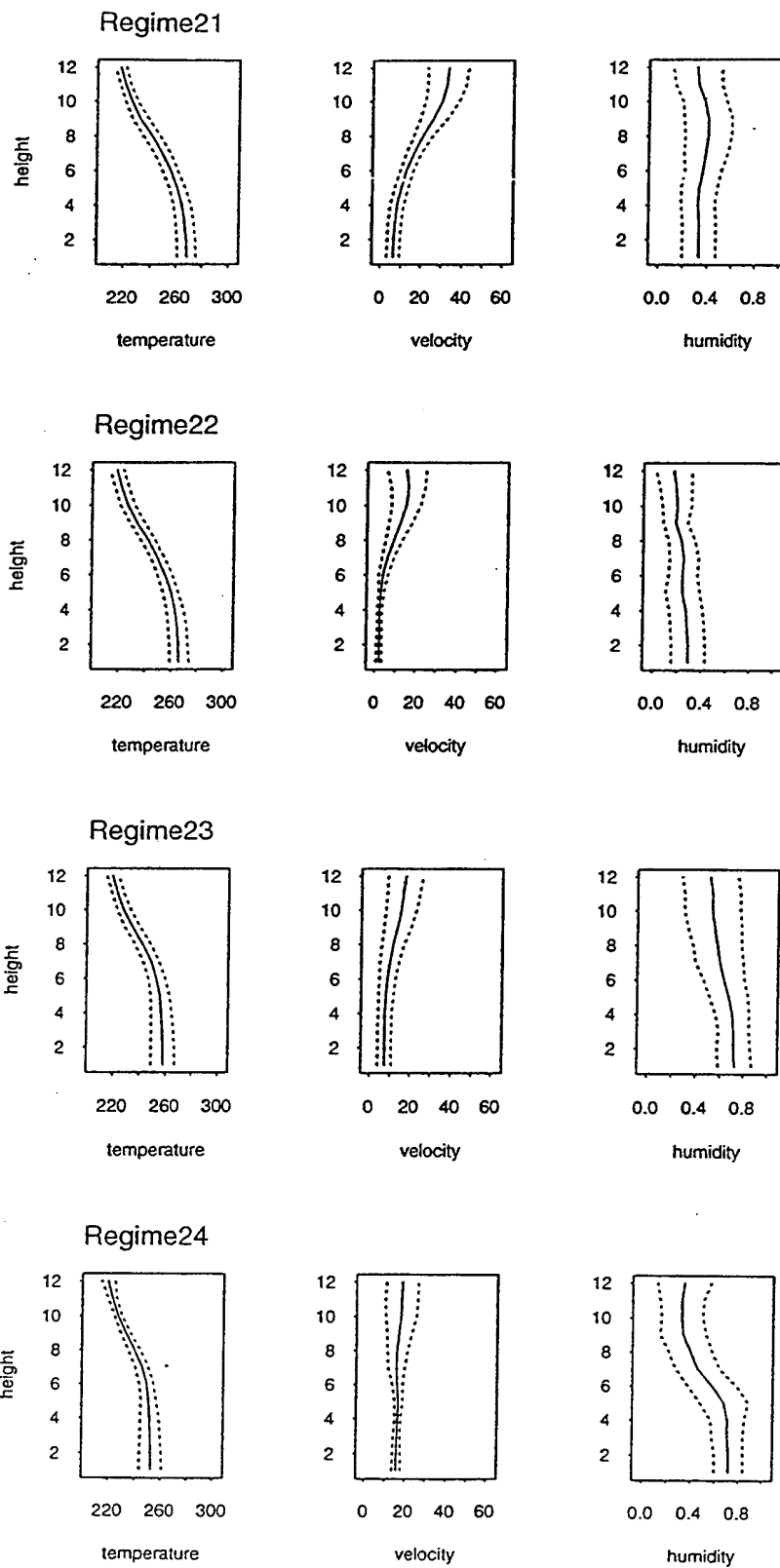


Figure 42: Continued.

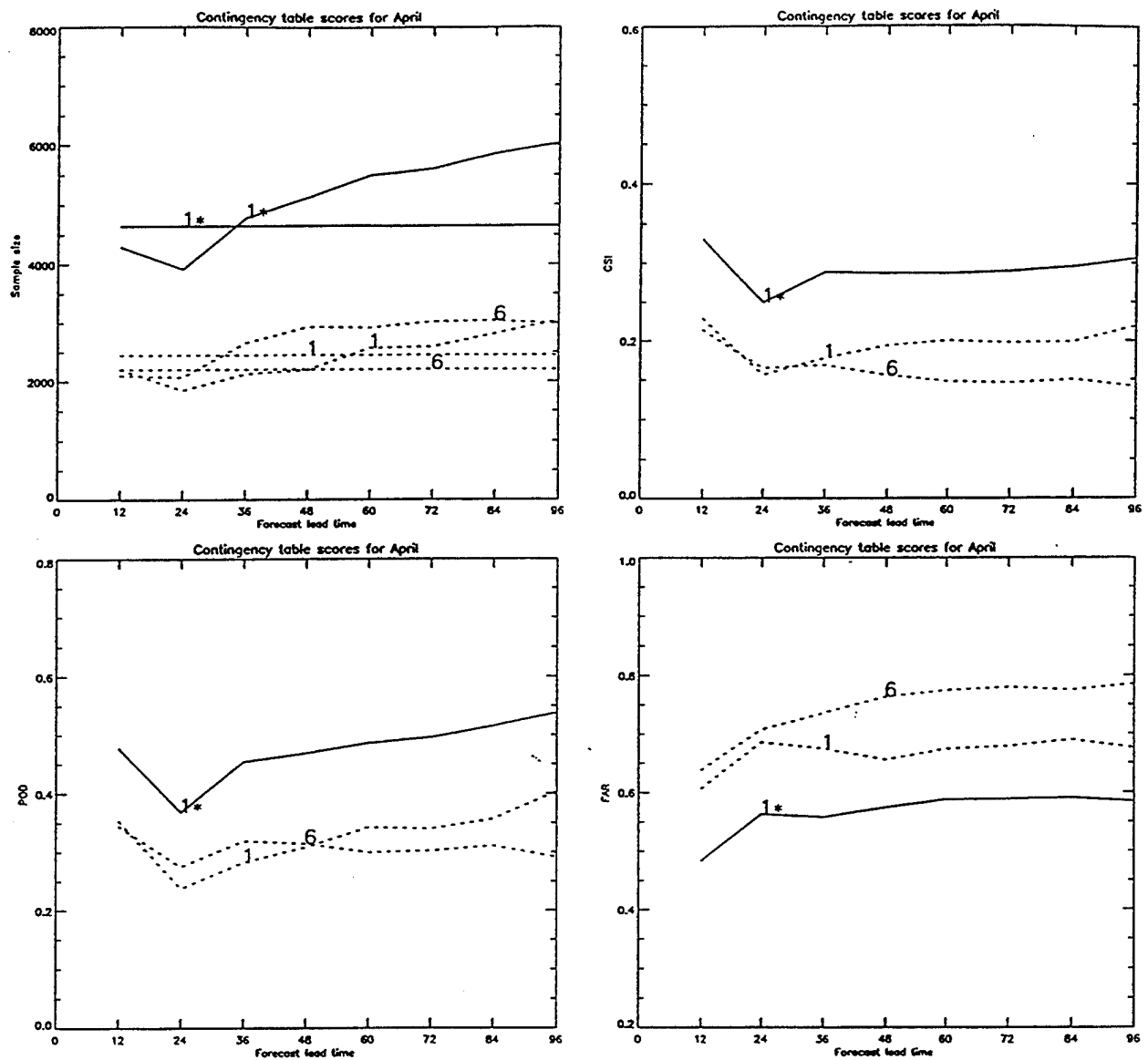


Figure 43: Sample sizes and contingency table scores of the combinations and individual regimes, for forecast lead times from 12 hours to 96 hours. Notice the change in scale for regime 23\*.

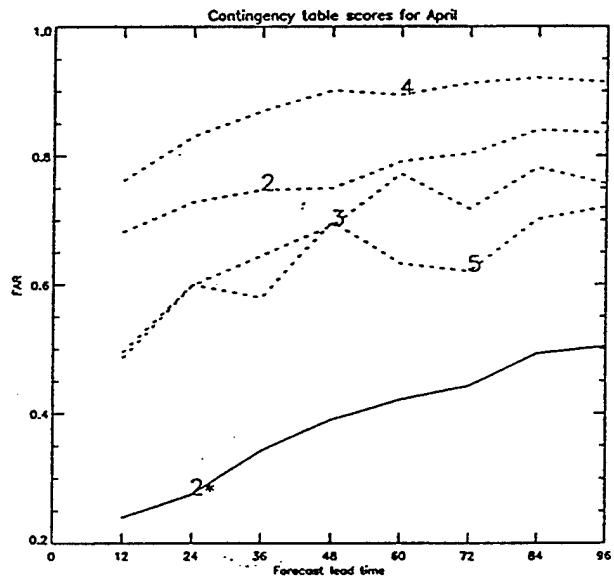
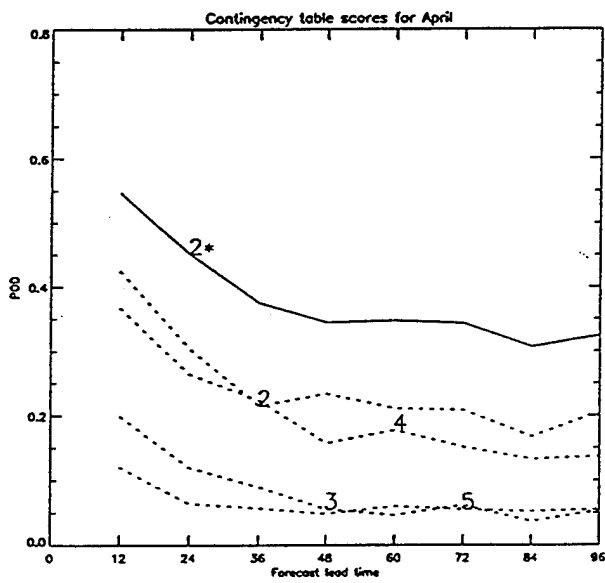
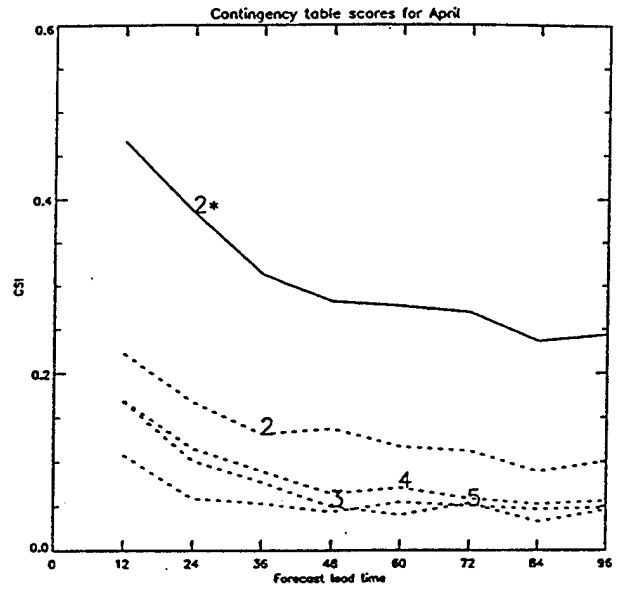
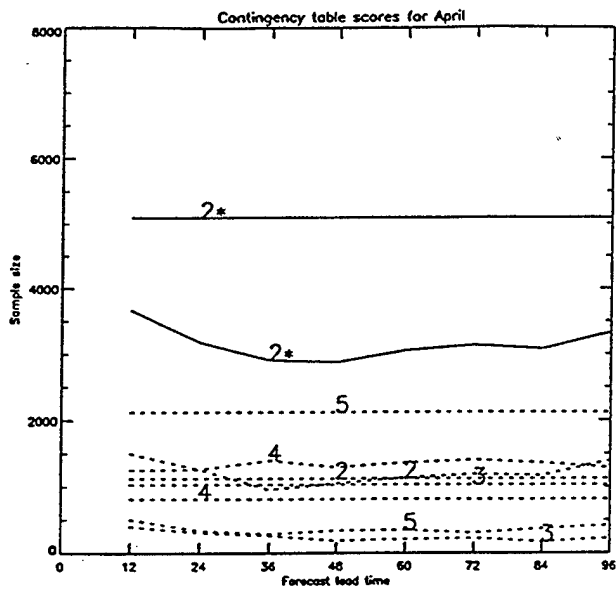


Figure 43: Continued.

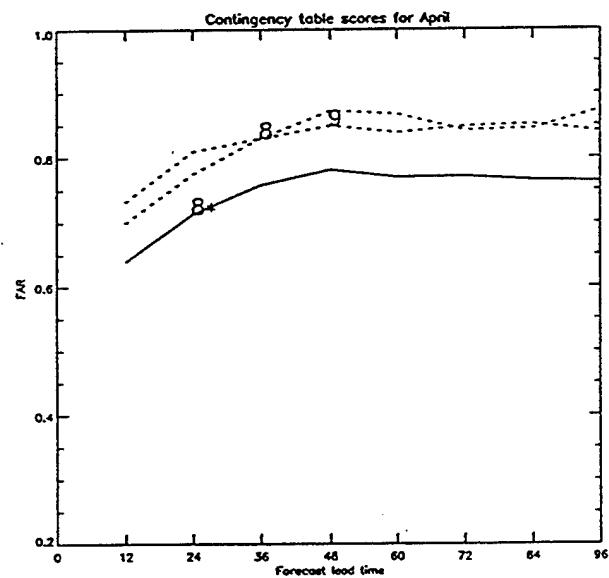
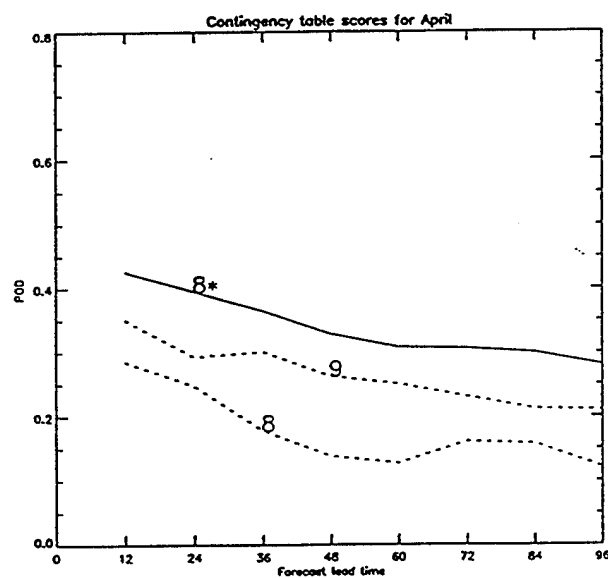
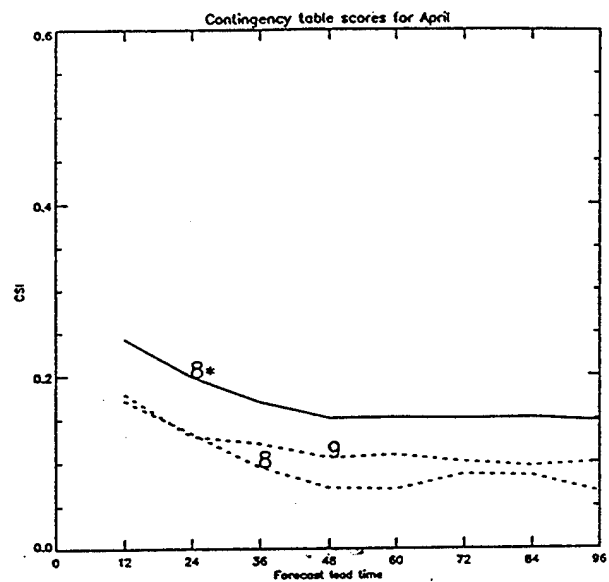
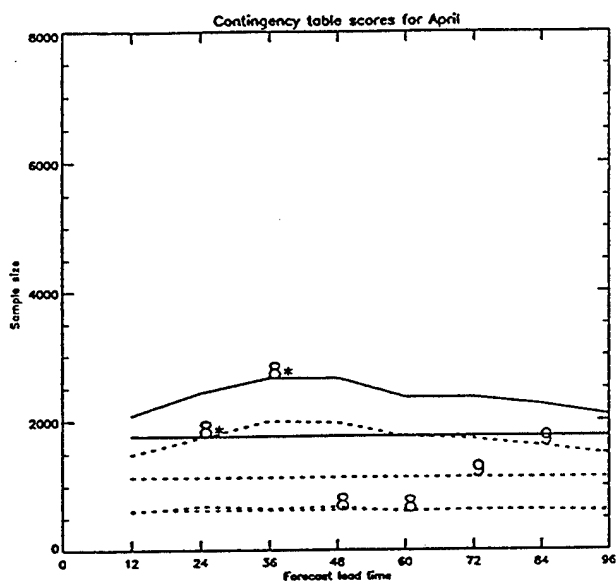


Figure 43: Continued.



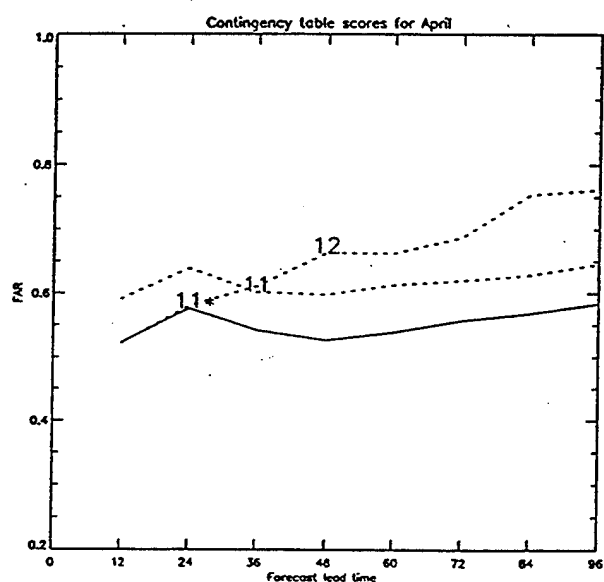
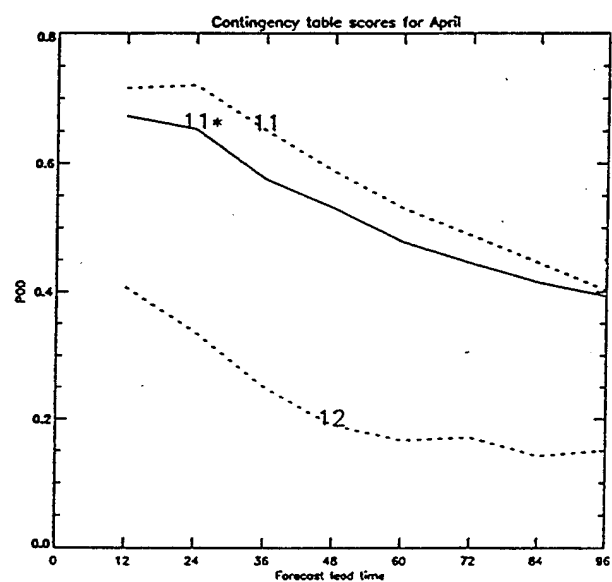
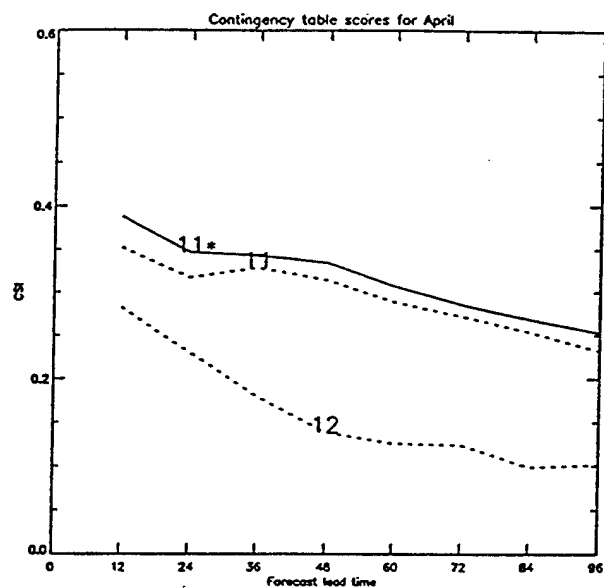
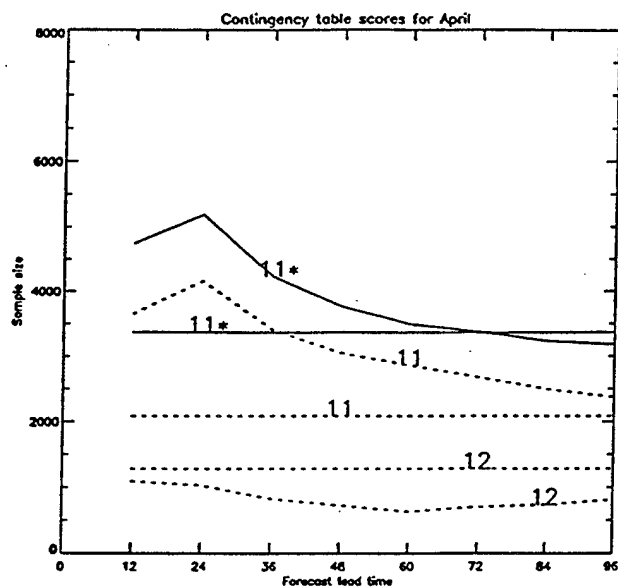


Figure 43: Continued.

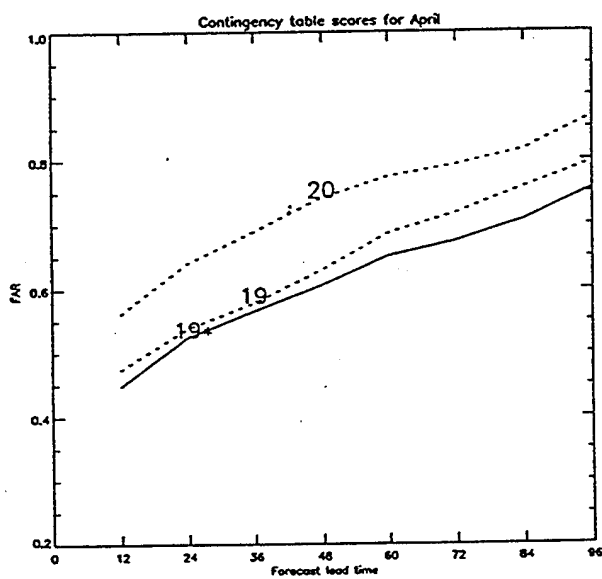
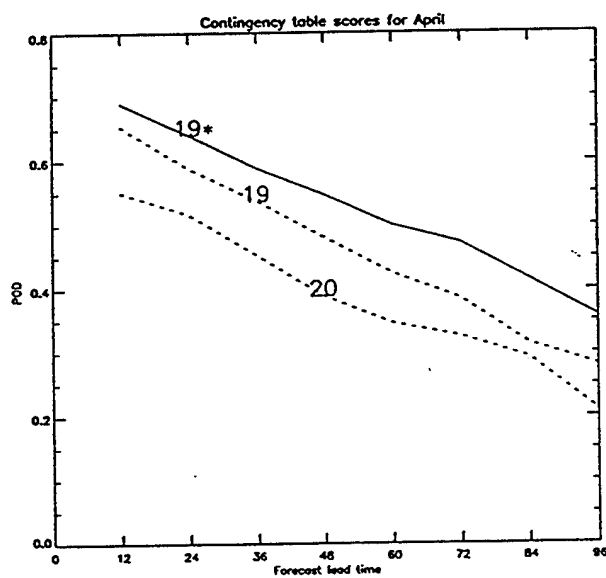
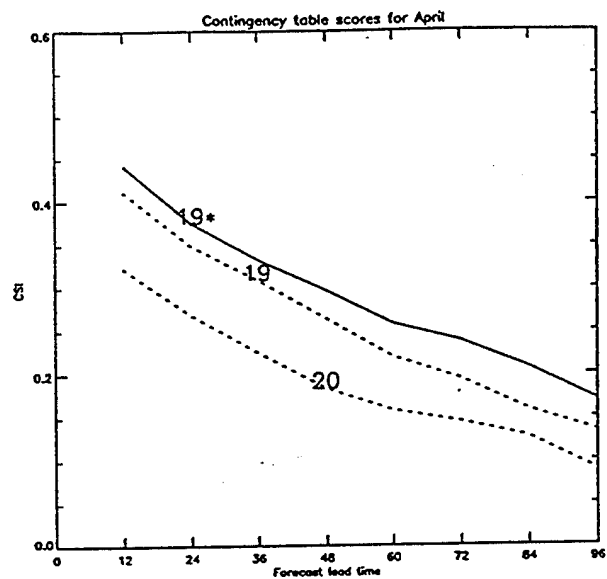
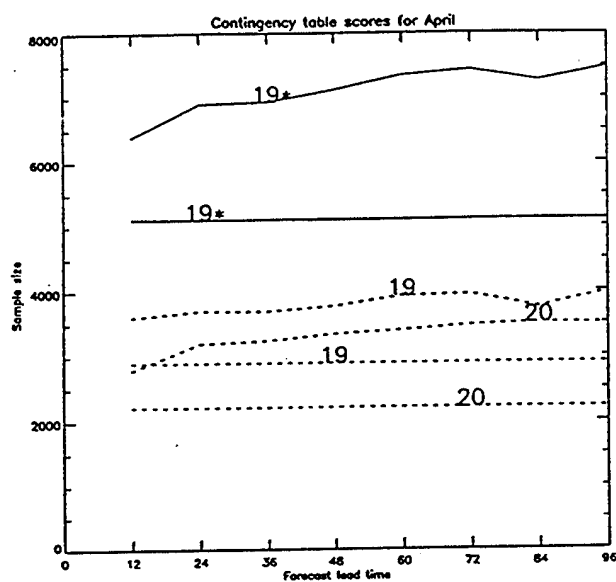


Figure 43: Continued.

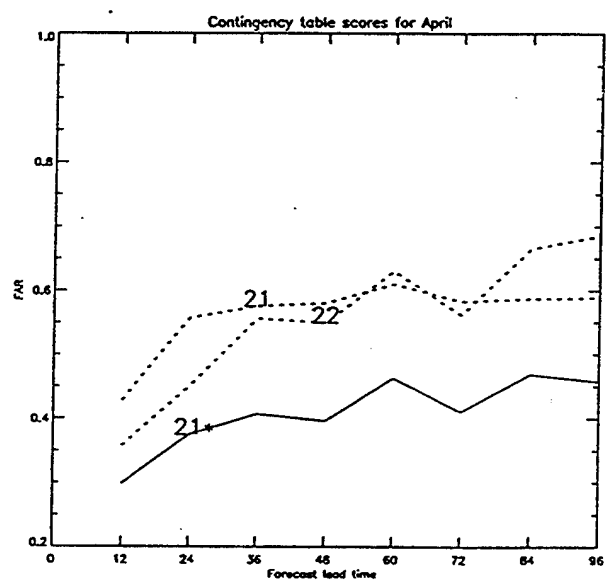
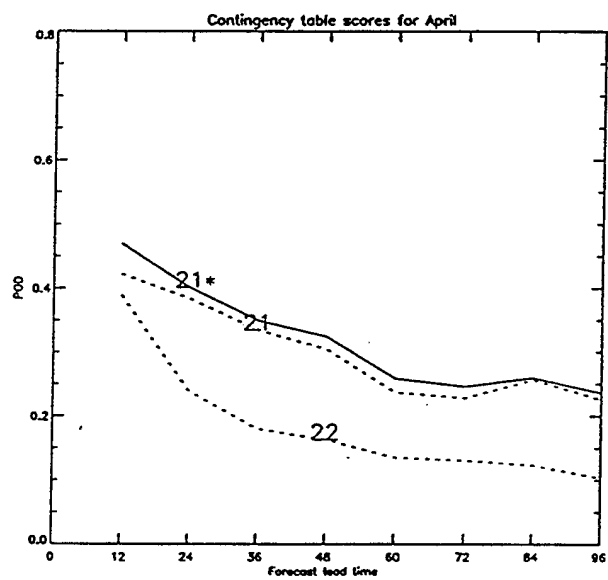
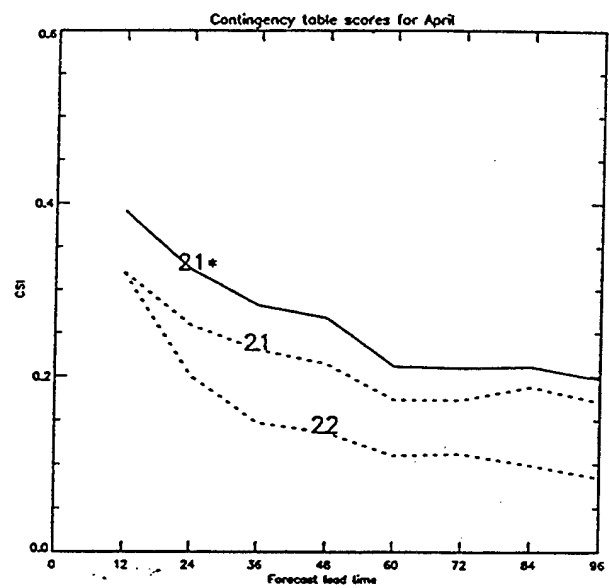
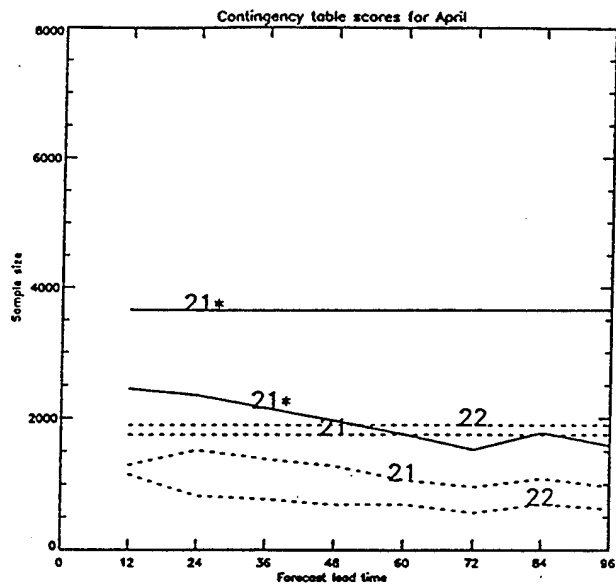


Figure 43: Continued.

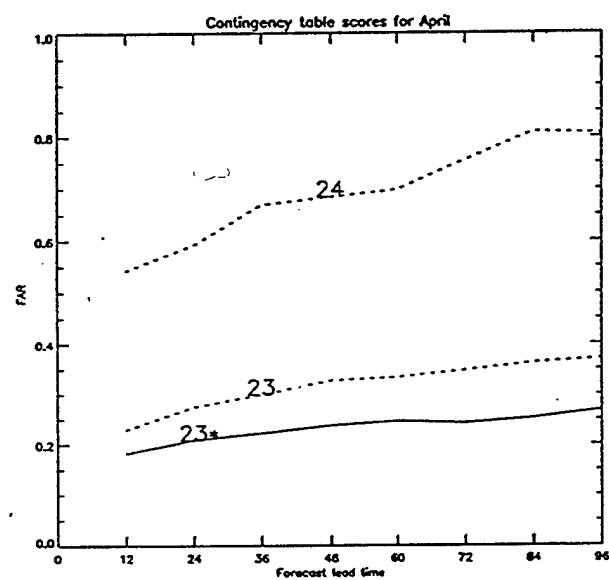
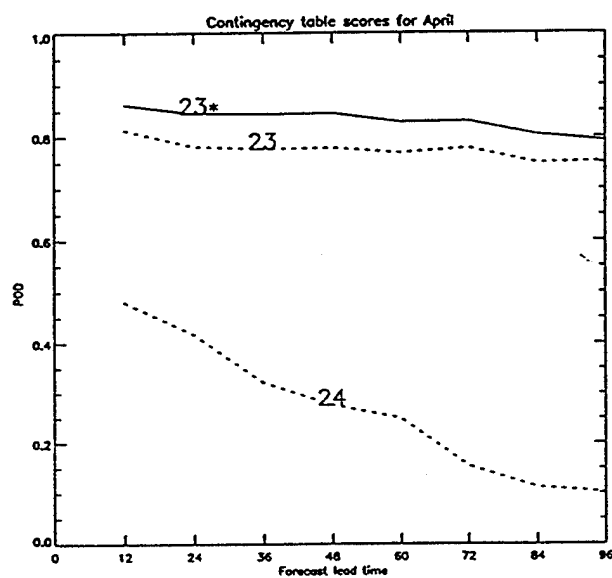
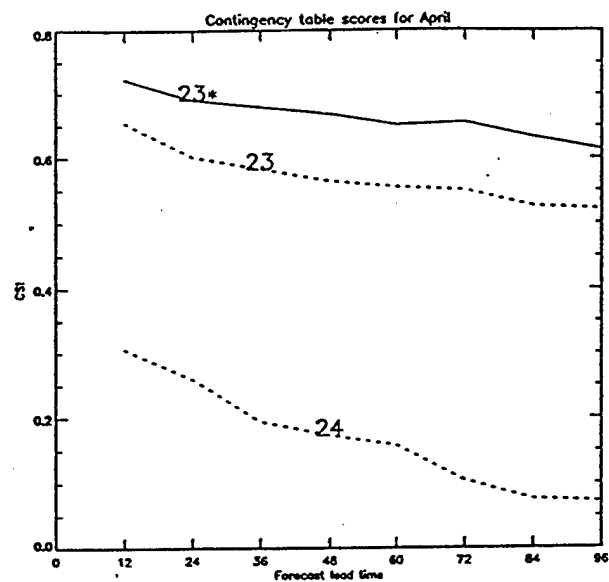
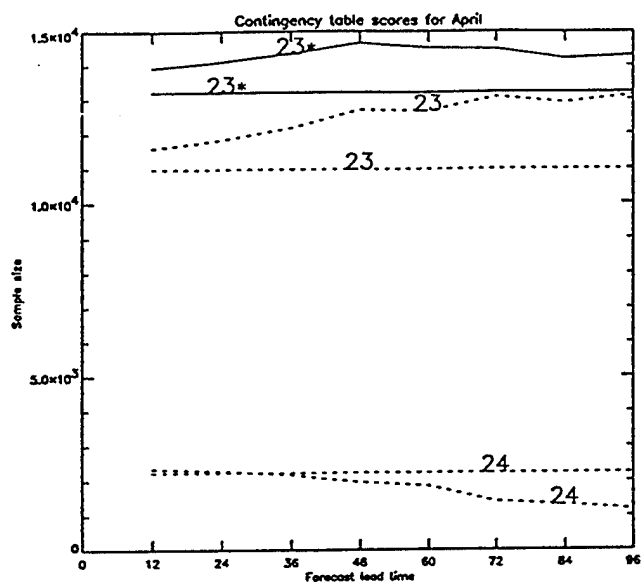


Figure 43: Continued.

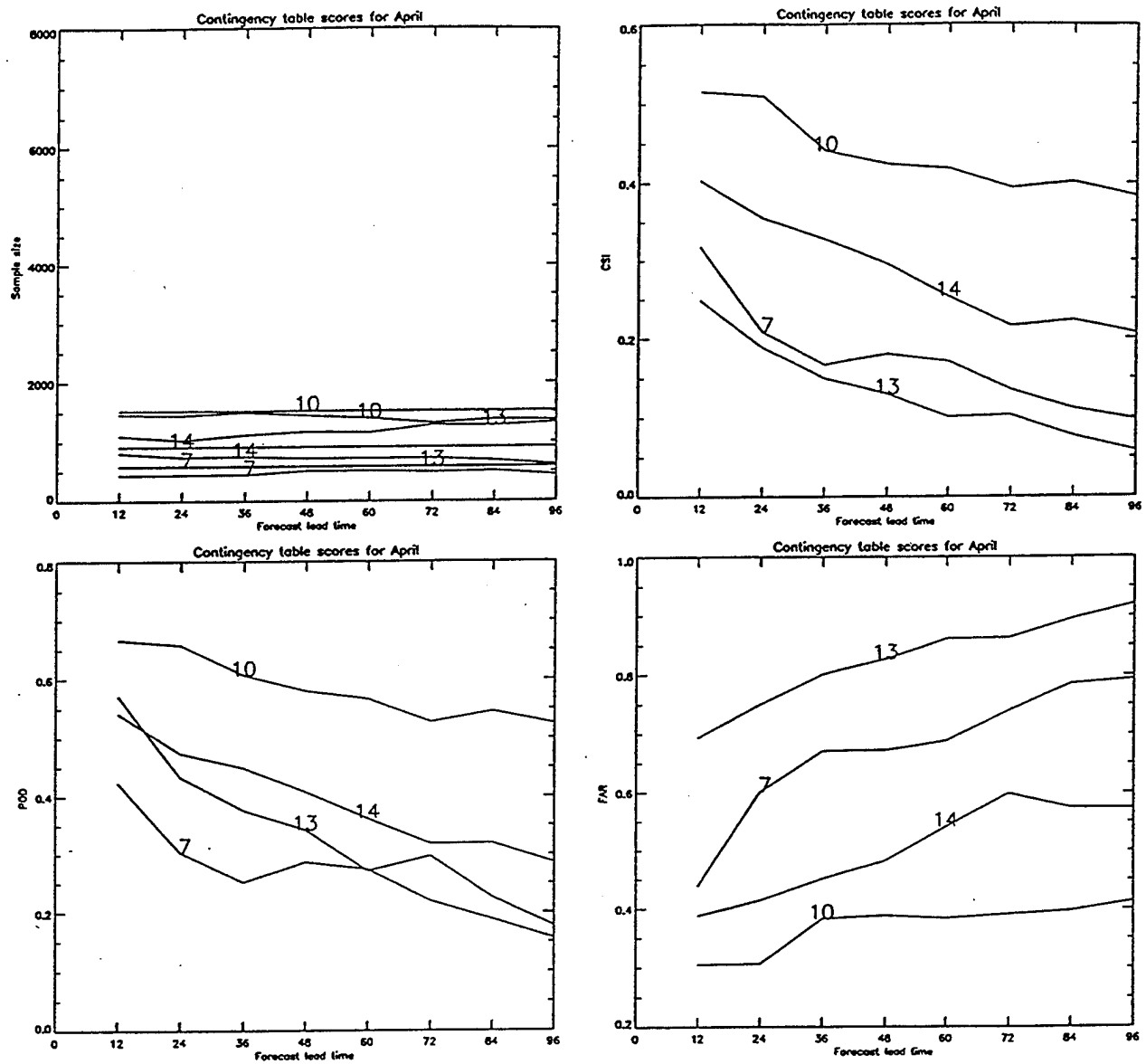


Figure 43: Continued.

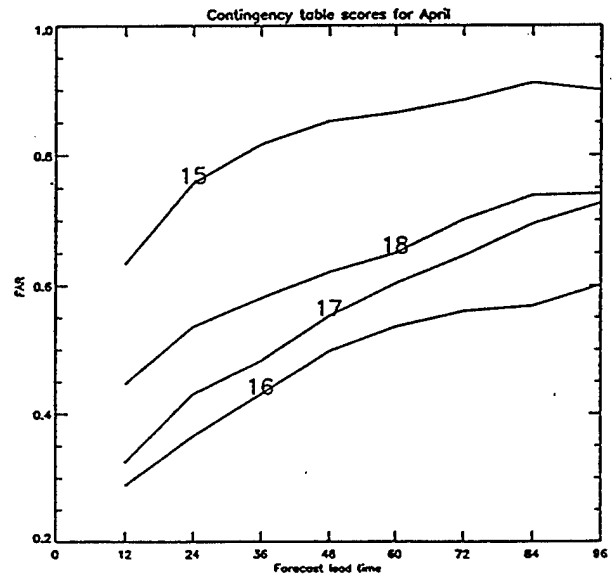
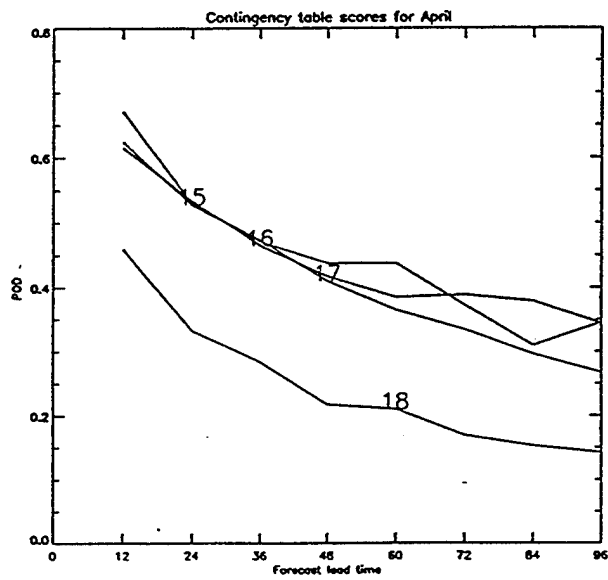
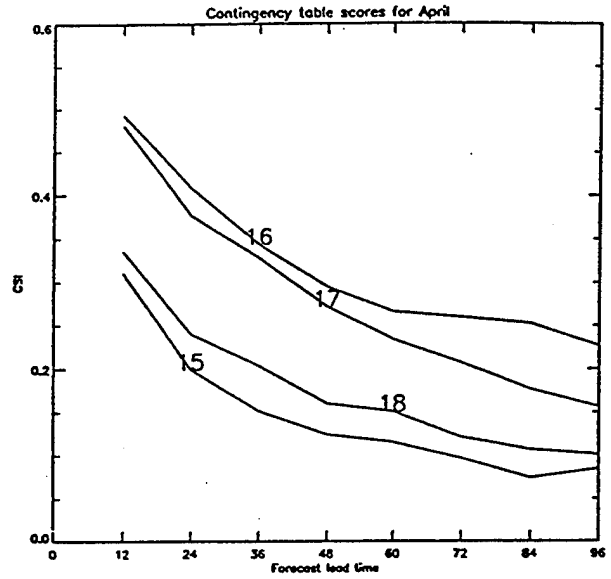
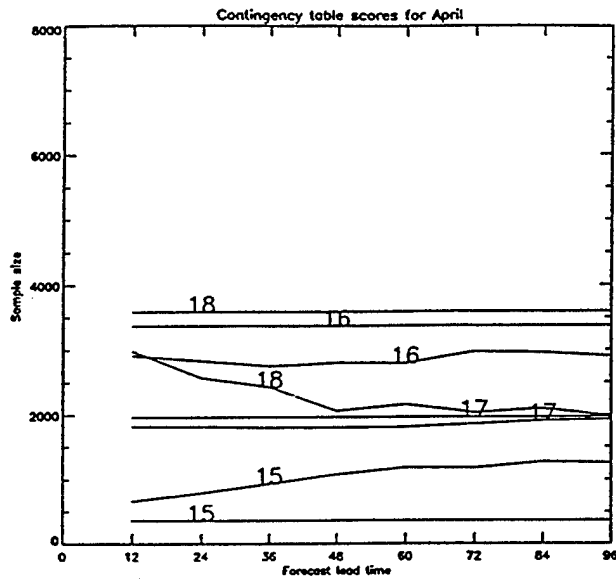


Figure 43: Continued.

Regimes and 1000–500 hPa thickness for 89/04/05/00 UTC

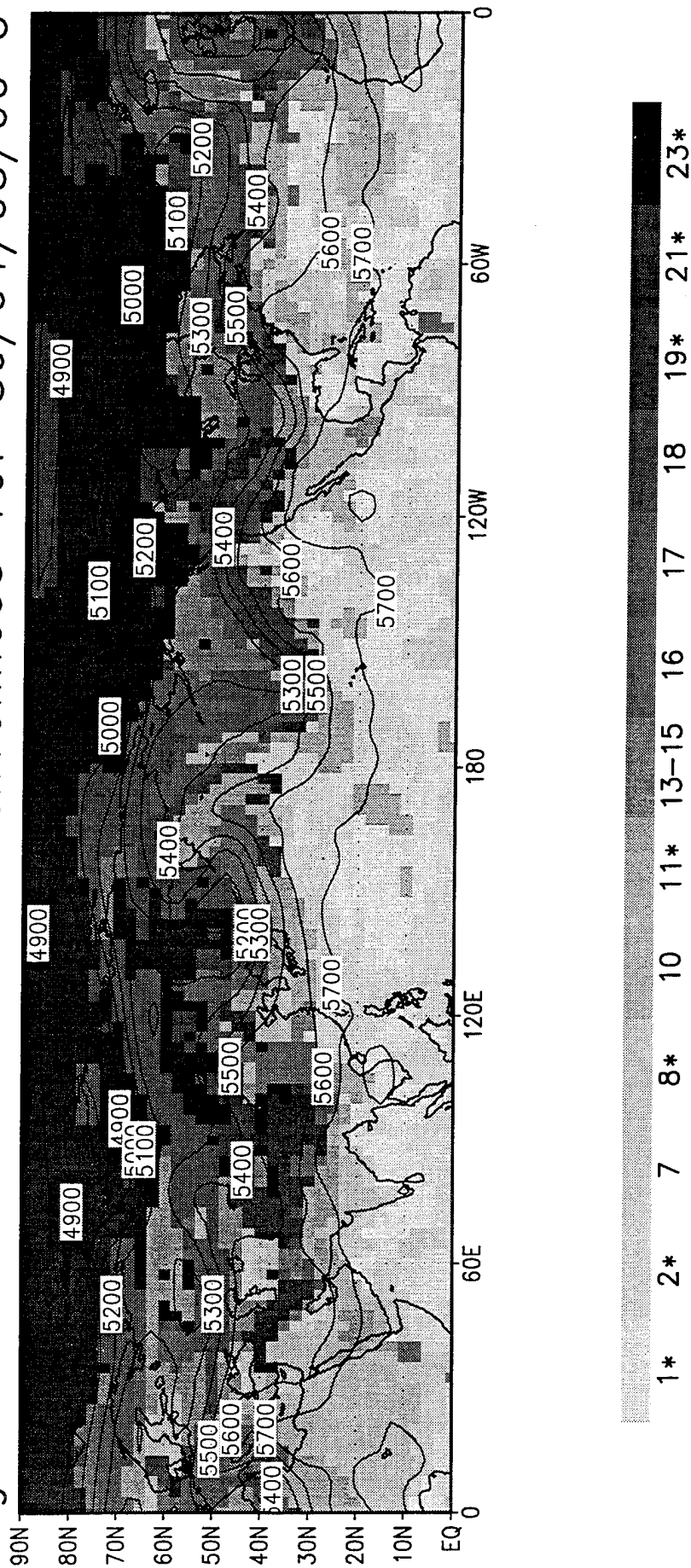


Figure 44: The geographic distribution of the regimes together with contours of the 1000hPa - 500 hPa thickness field for 00 UTC 5 April 1989.

# Regimes and 300 hPa wind speed for 89/04/05/00 UTC

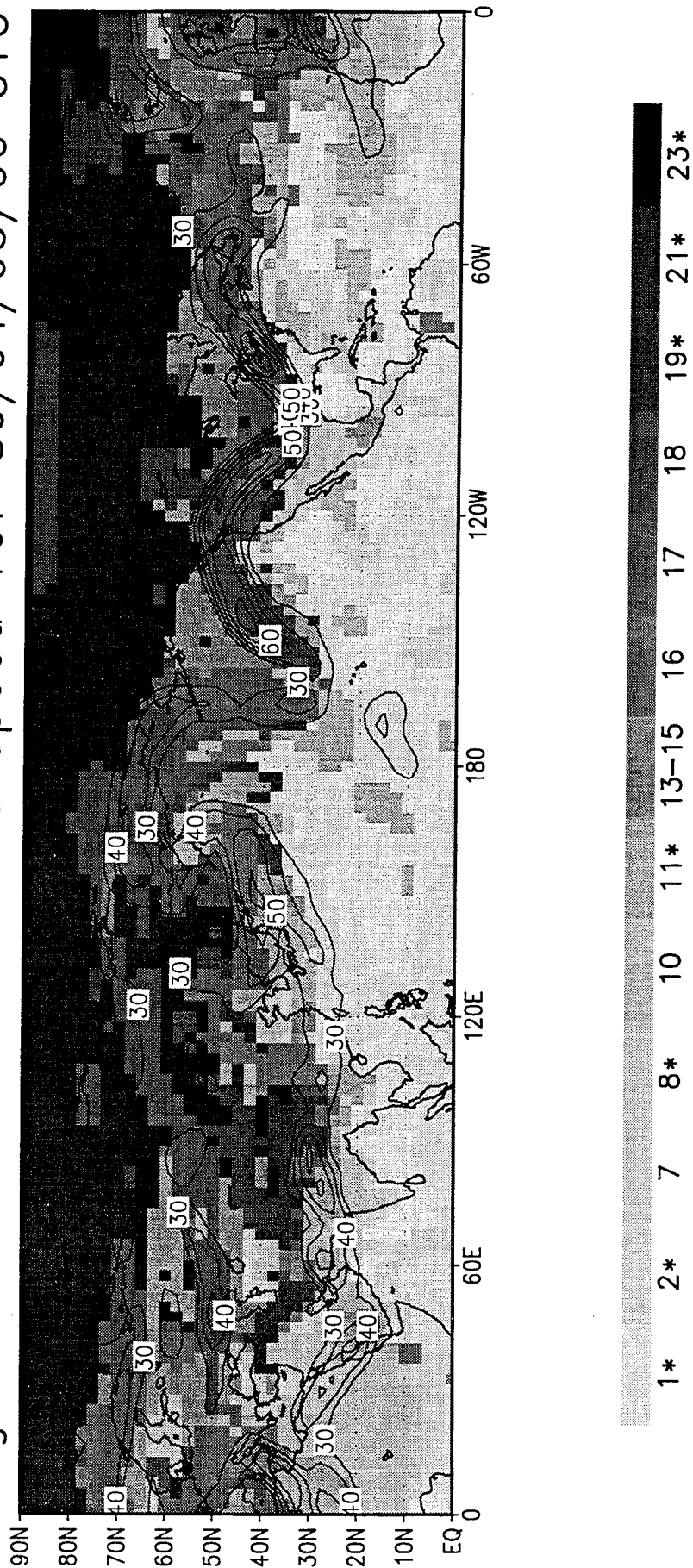


Figure 45: The geographic distribution of the regimes together with contours of the 300 hPa wind speed (every 10 m/s, only contours above 30 m/s shown) for 00 UTC 5 April 1989.



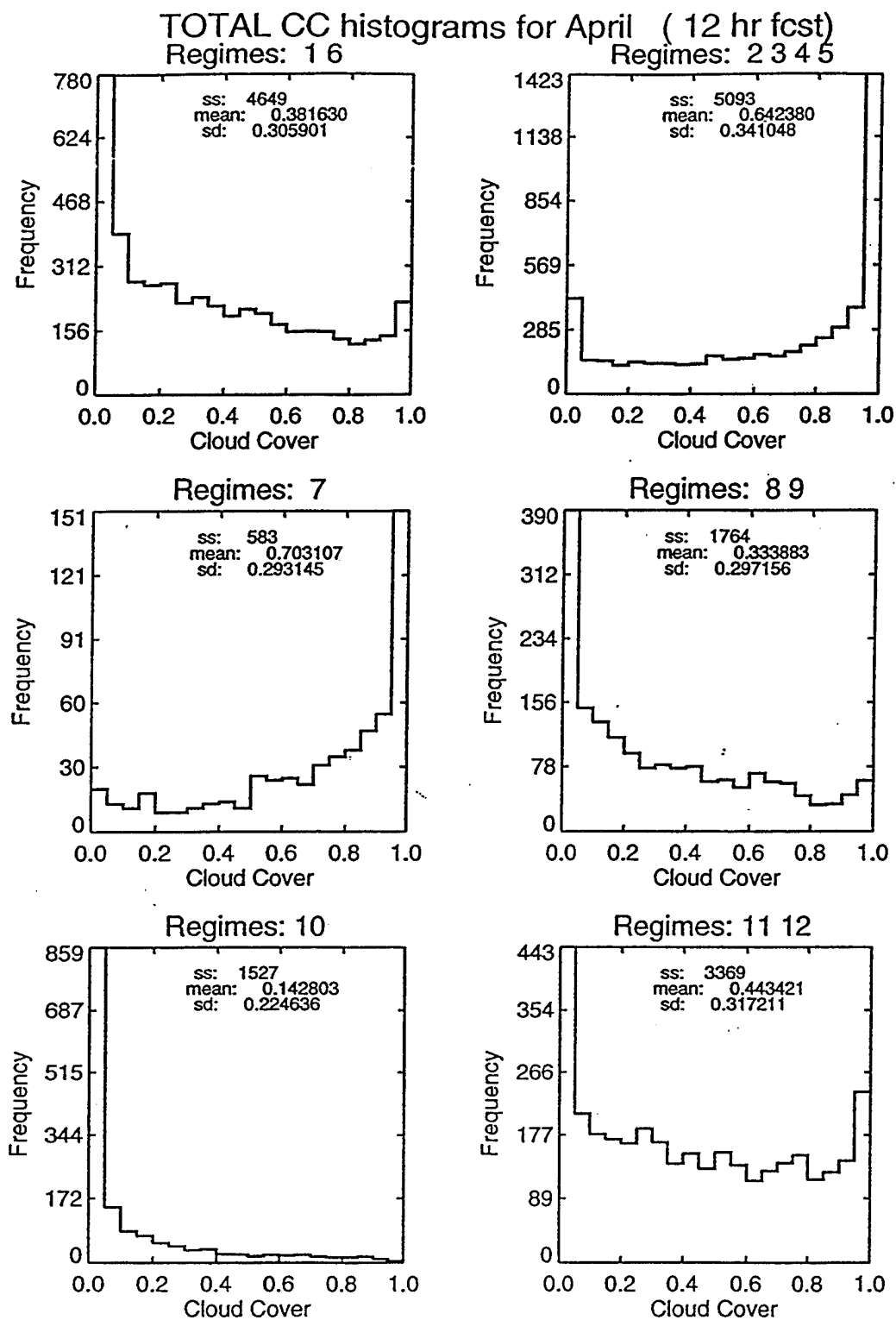


Figure 46: Histograms of total cloud cover for a one-week period in April for the 15 regimes.

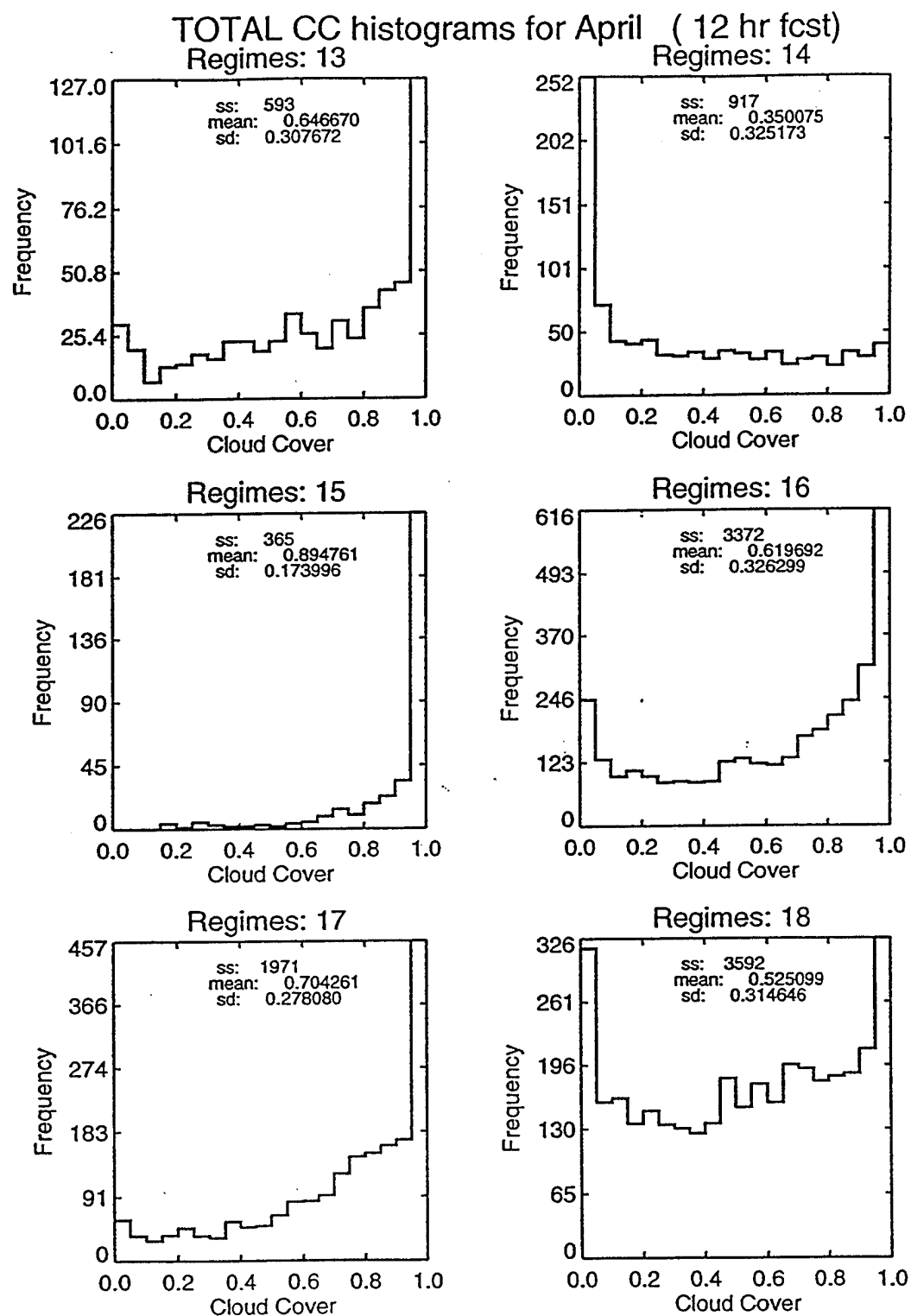


Figure 46: Continued

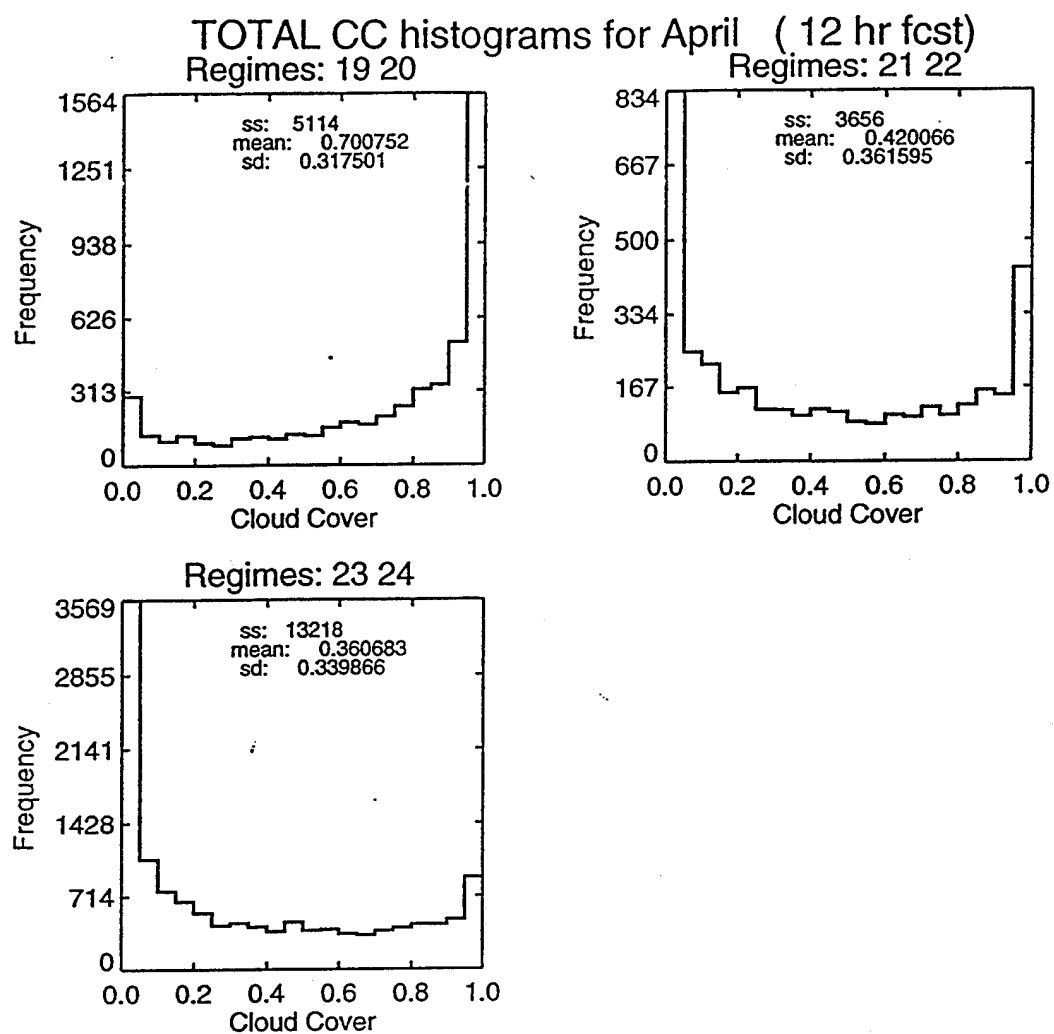


Figure 46: Continued.

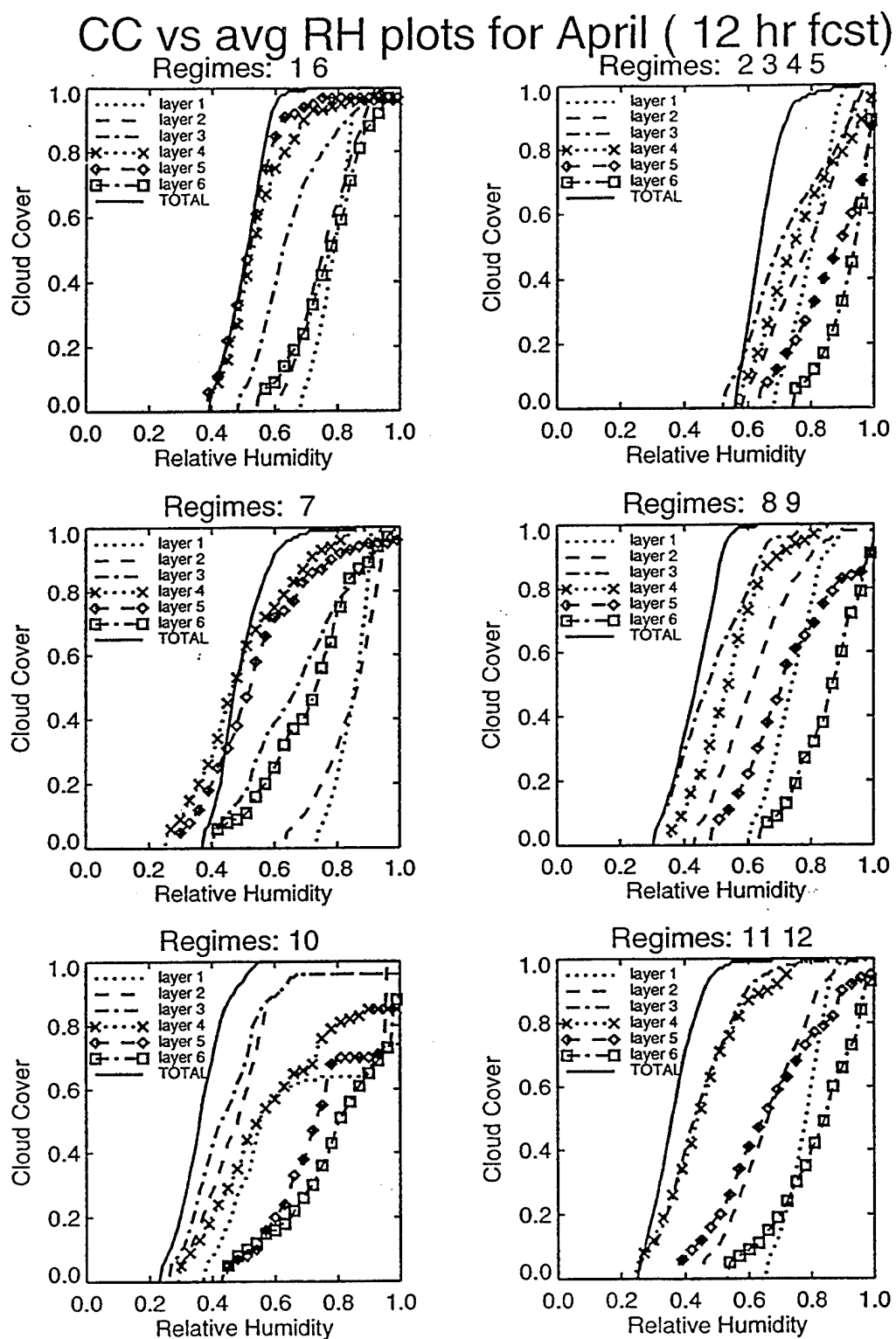


Figure 47: CCA curves for 12 hour forecasts, using the average RH of the cloud layers, for the 15 regimes.

# CC vs avg RH plots for April ( 12 hr fcst)

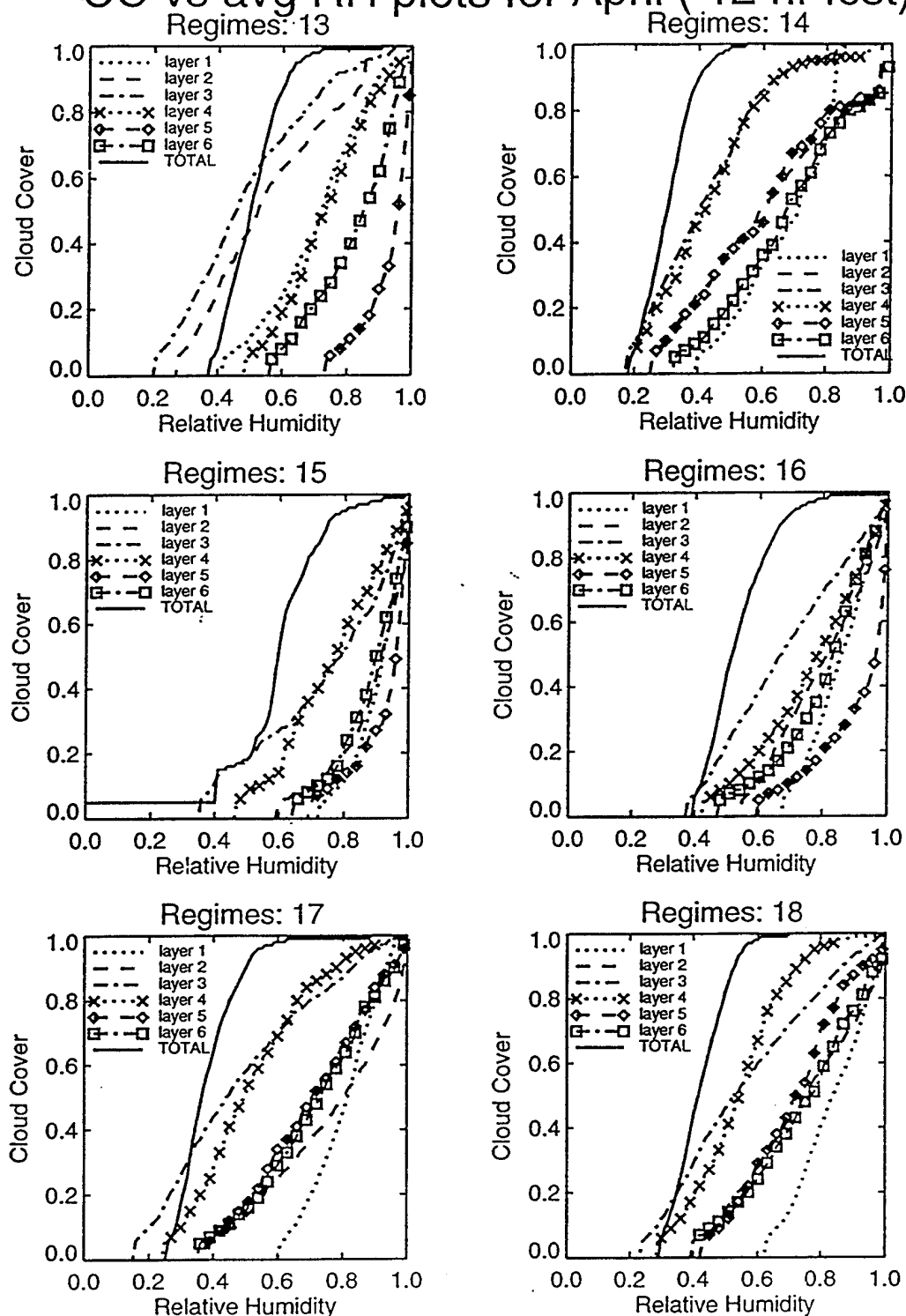


Figure 47: Continued.

# CC vs avg RH plots for April ( 12 hr fcst)

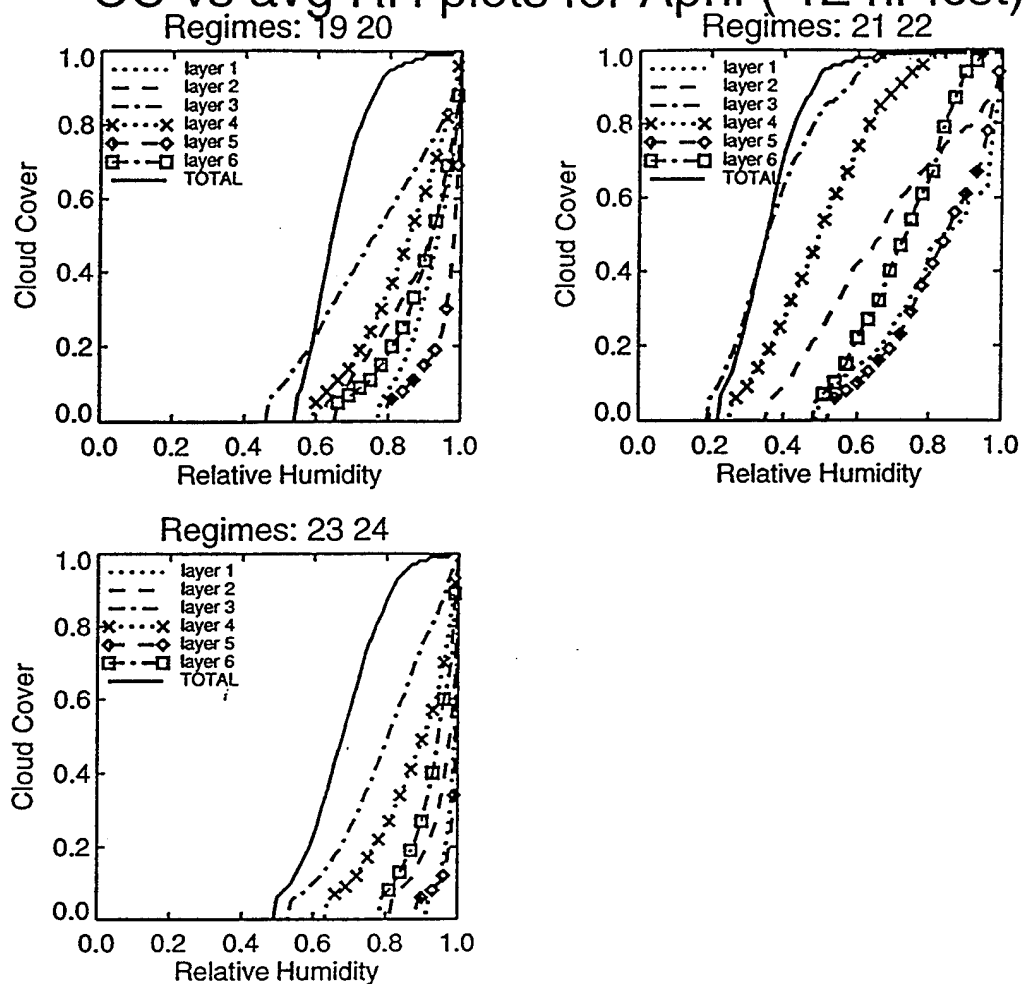


Figure 47: Continued.

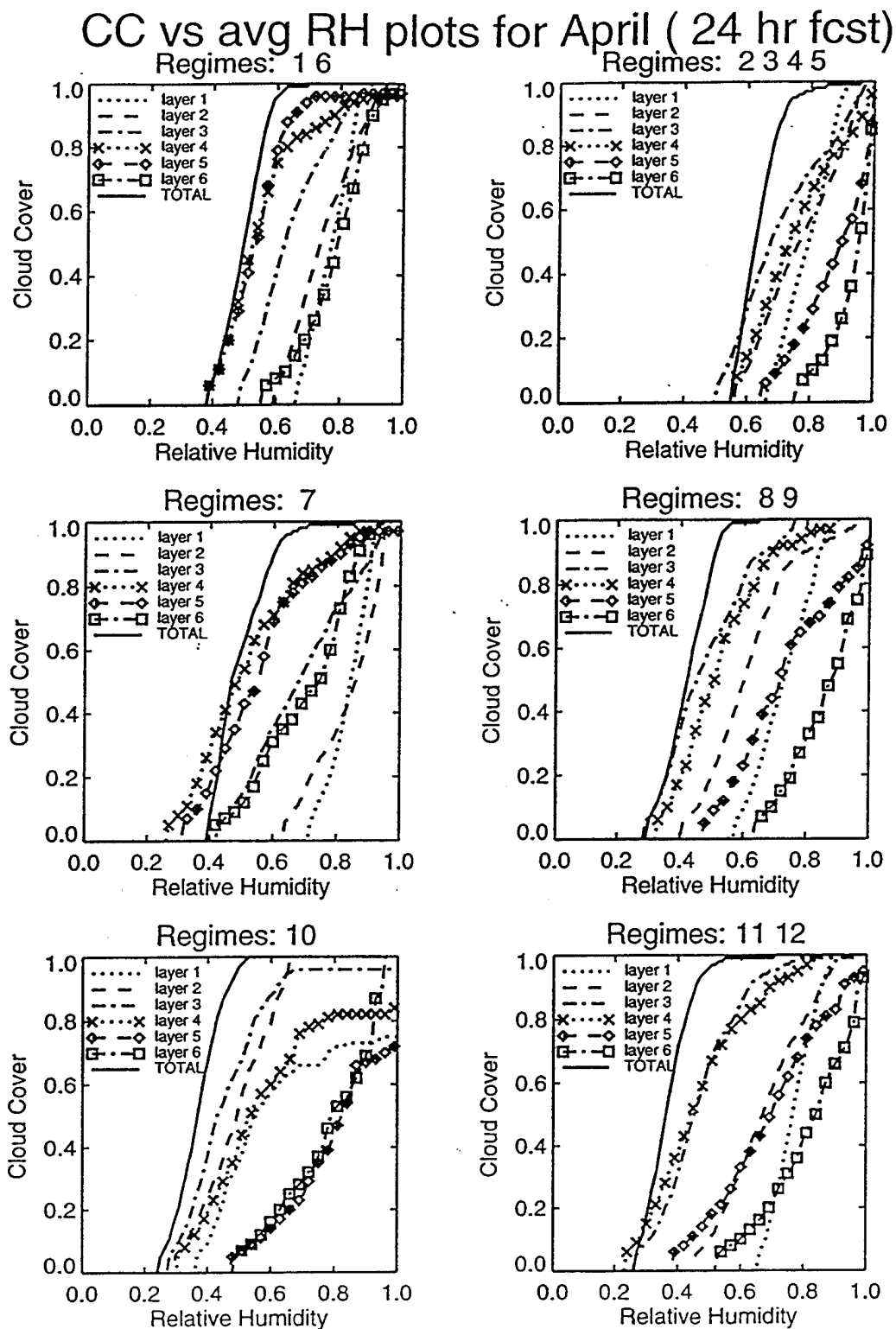


Figure 48: CCA curves for 24 hour forecasts, using the average RH of the cloud layers, for the 15 regimes.

# CC vs avg RH plots for April ( 24 hr fcst)

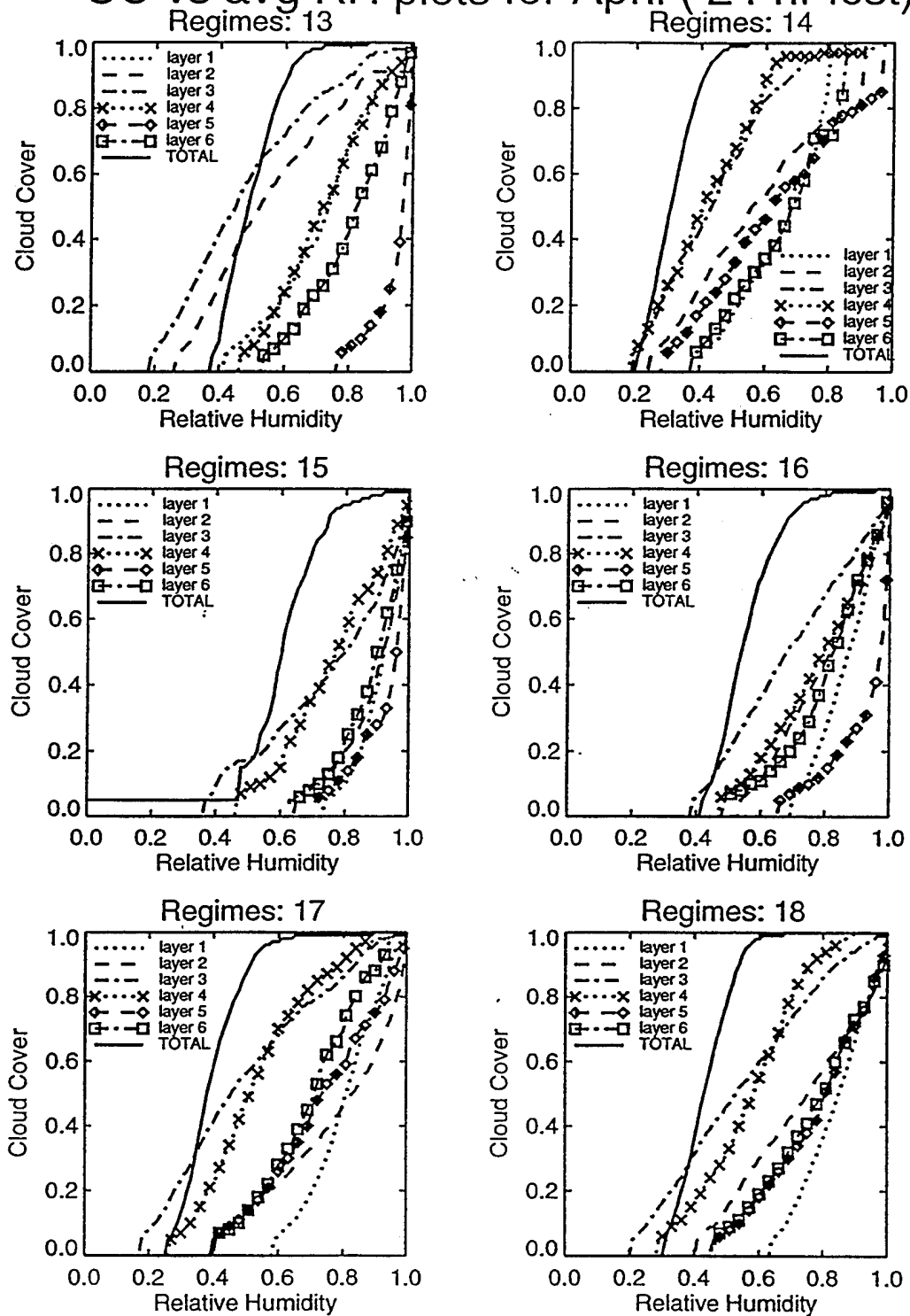


Figure 48: Continued.



# CC vs avg RH plots for April ( 24 hr fcst)

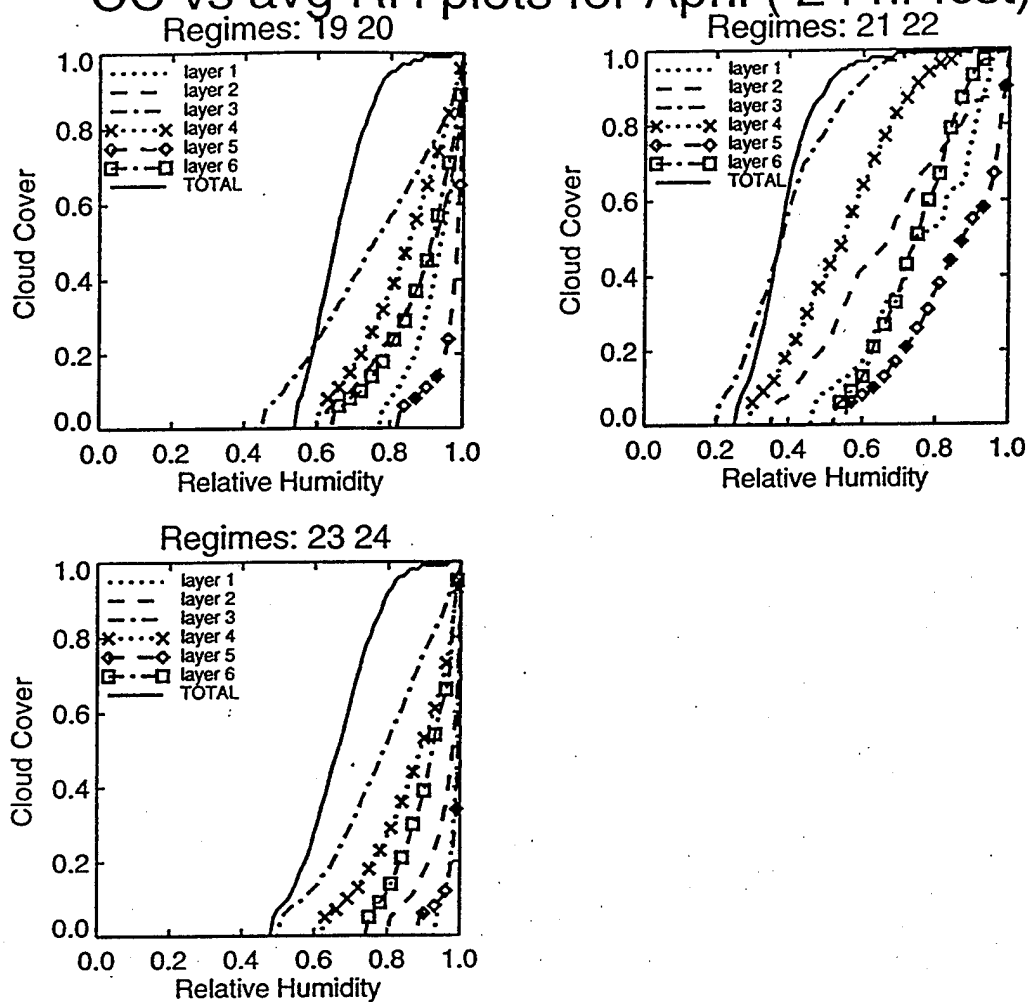


Figure 48: Continued.

## Octagon, fronthalf, and backhalf

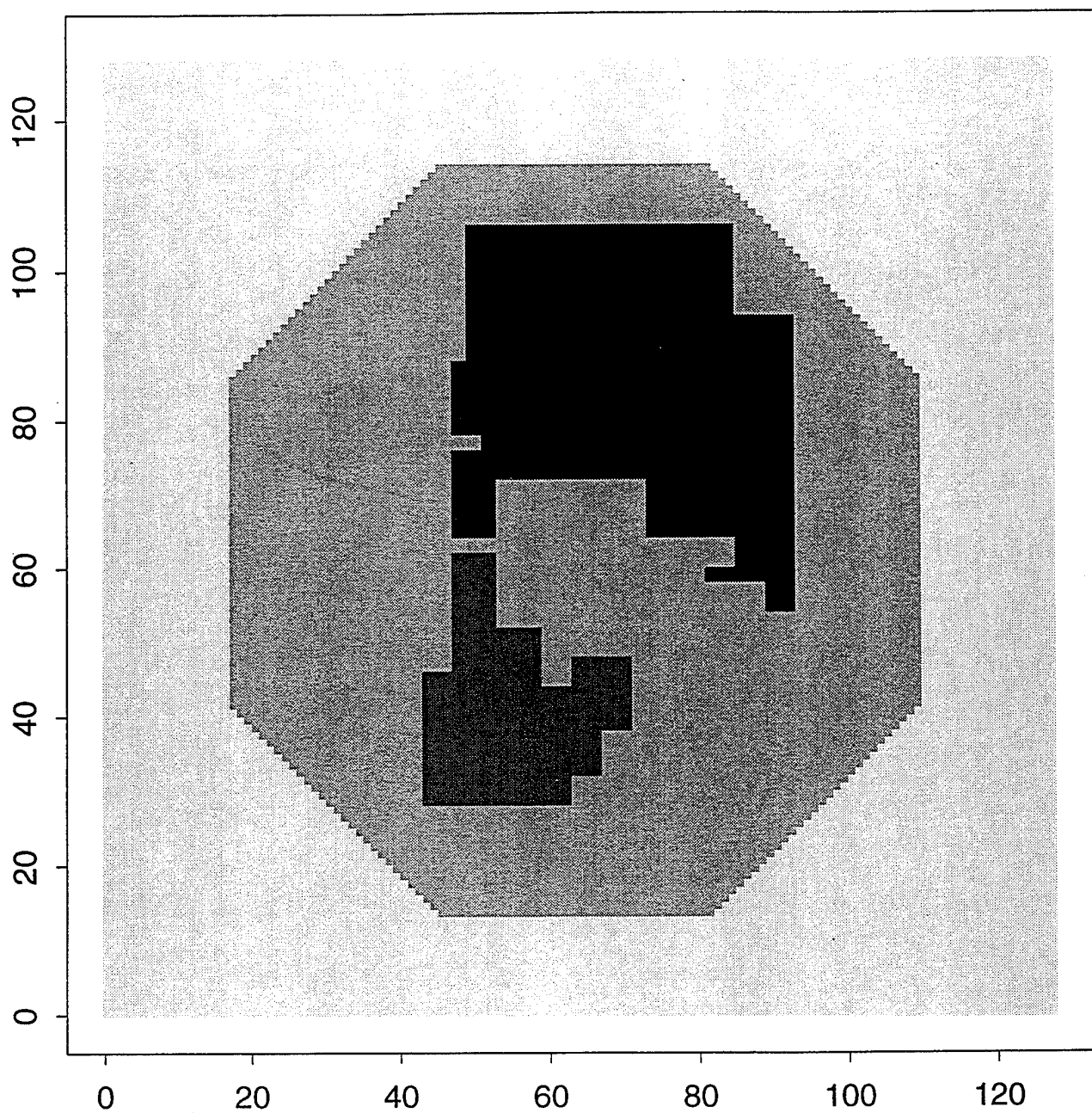


Figure 49: The Northern Hemisphere octagon, shown in light shading over the RTNEPH grid. Axis labels are the distance from the lower left corner of the grid in units of 1/2-mesh grid lengths. Also shown in darker shading are areas encompassing the Eurasian and North American continents, denoted in the verification statistics as "backhalf" and "fronthalf", respectively.

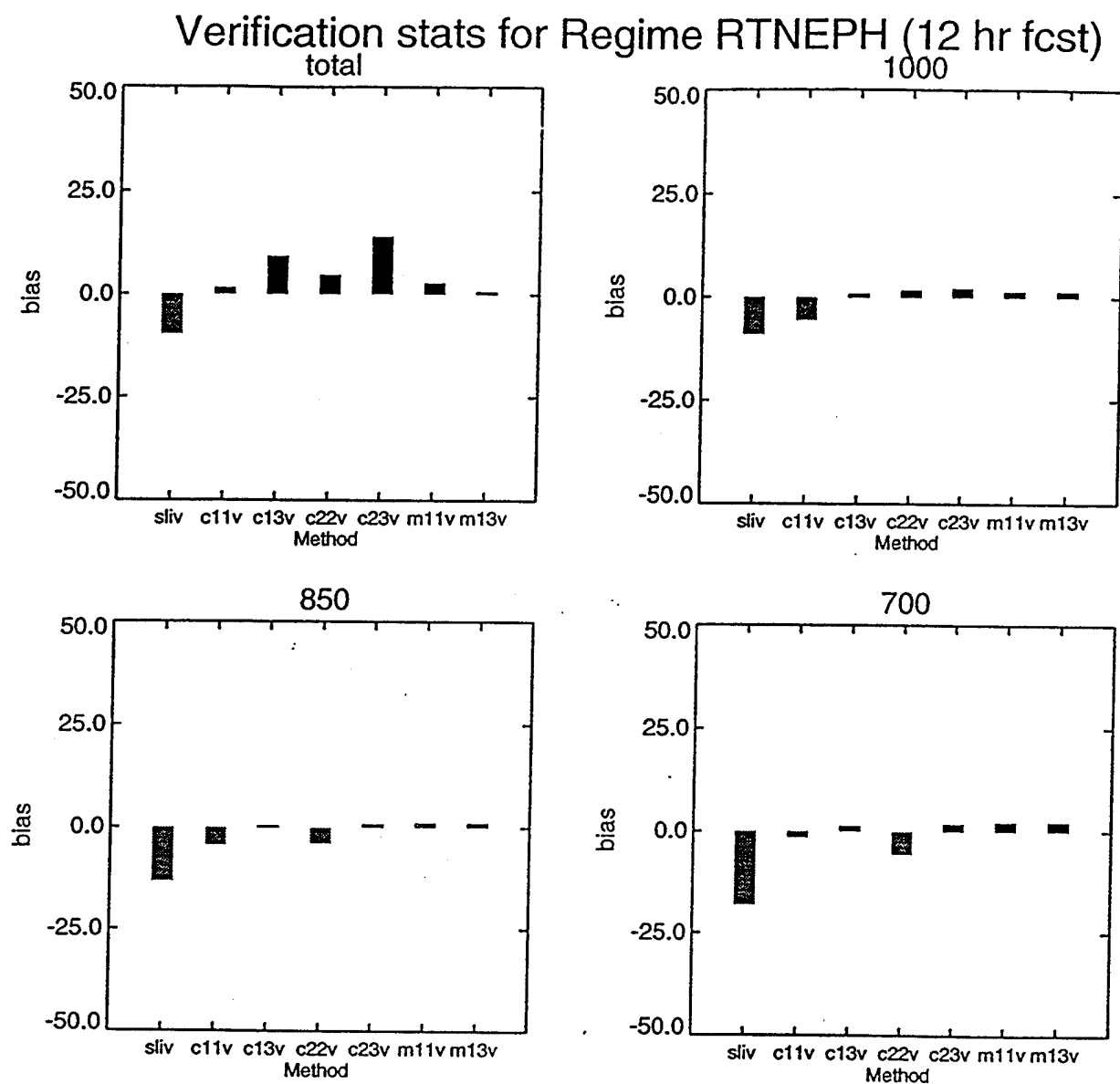


Figure 50: Bias of 12-hour forecasts of cloud cover (Total and the six RTNEPH layers) of the candidate cloud cover schemes for the dependent data set. Slingo scheme is labeled sli, average cloud cover scheme with (without) overlap computation for total cloud cover is labeled m13 (m11), CCA scheme using average RH and separate total cloudcover curve (overlap computation) is labeled c11 (c13), corresponding maximum RH CCA schemes are labeled c22 and c23.

# Verification stats for Regime RTNEPH (12 hr fcst)

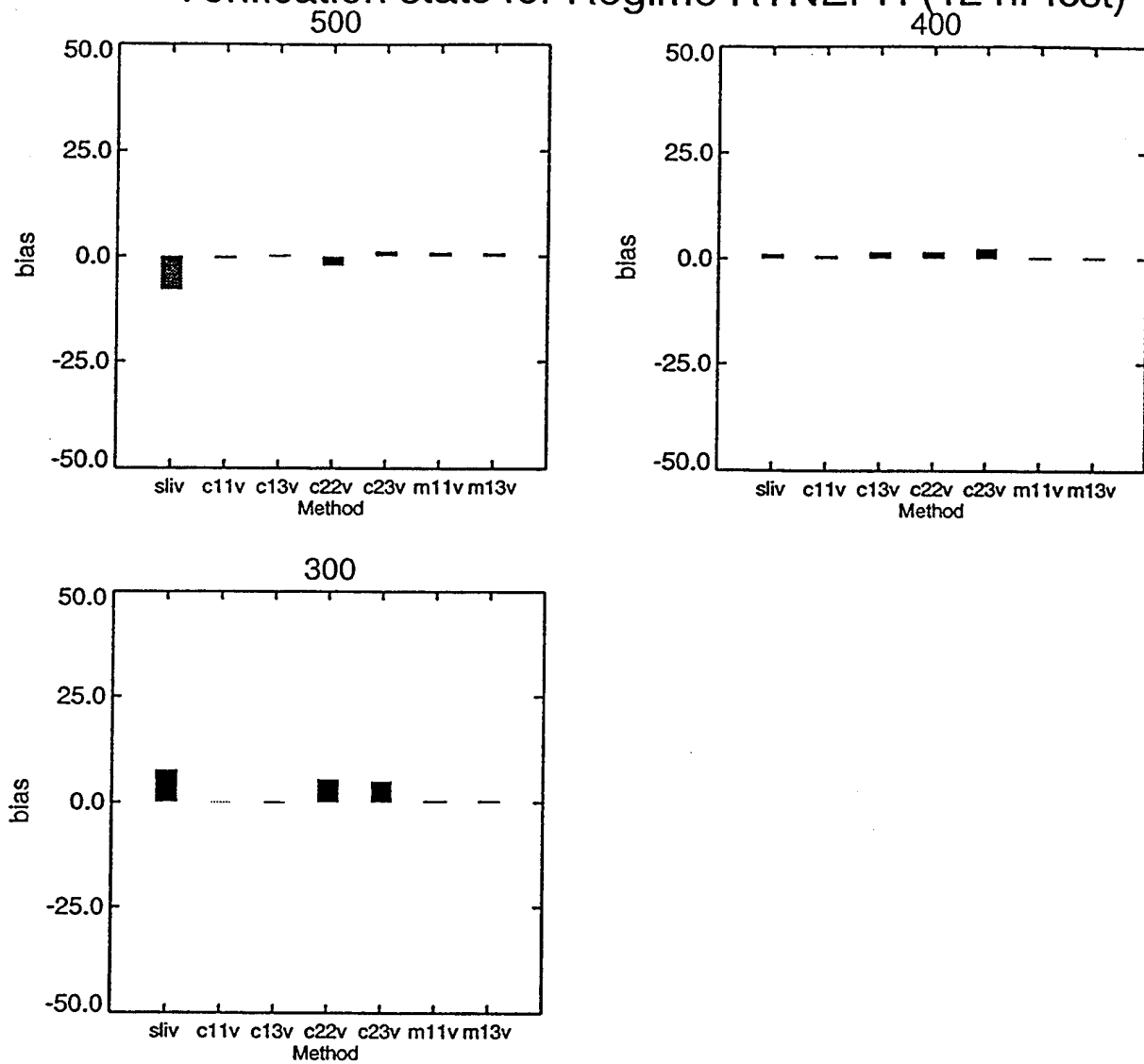


Figure 50: Continued.

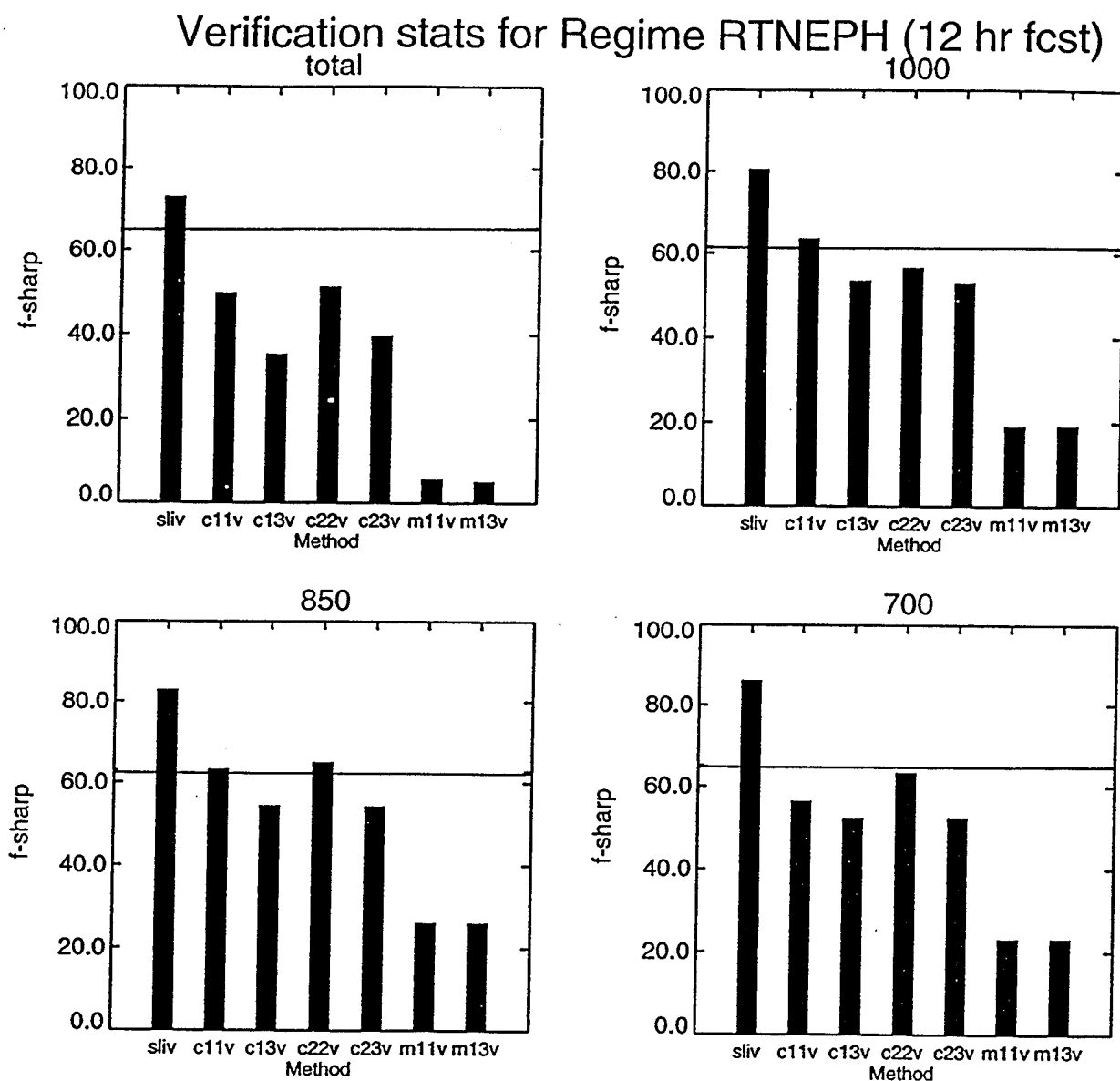


Figure 51: As Figure 50, except for the sharpness of 12-hour cloud cover forecasts. Sharpness of the verifying RTNEPH analyses is indicated by horizontal line.

# Verification stats for Regime RTNEPH (12 hr fcst)

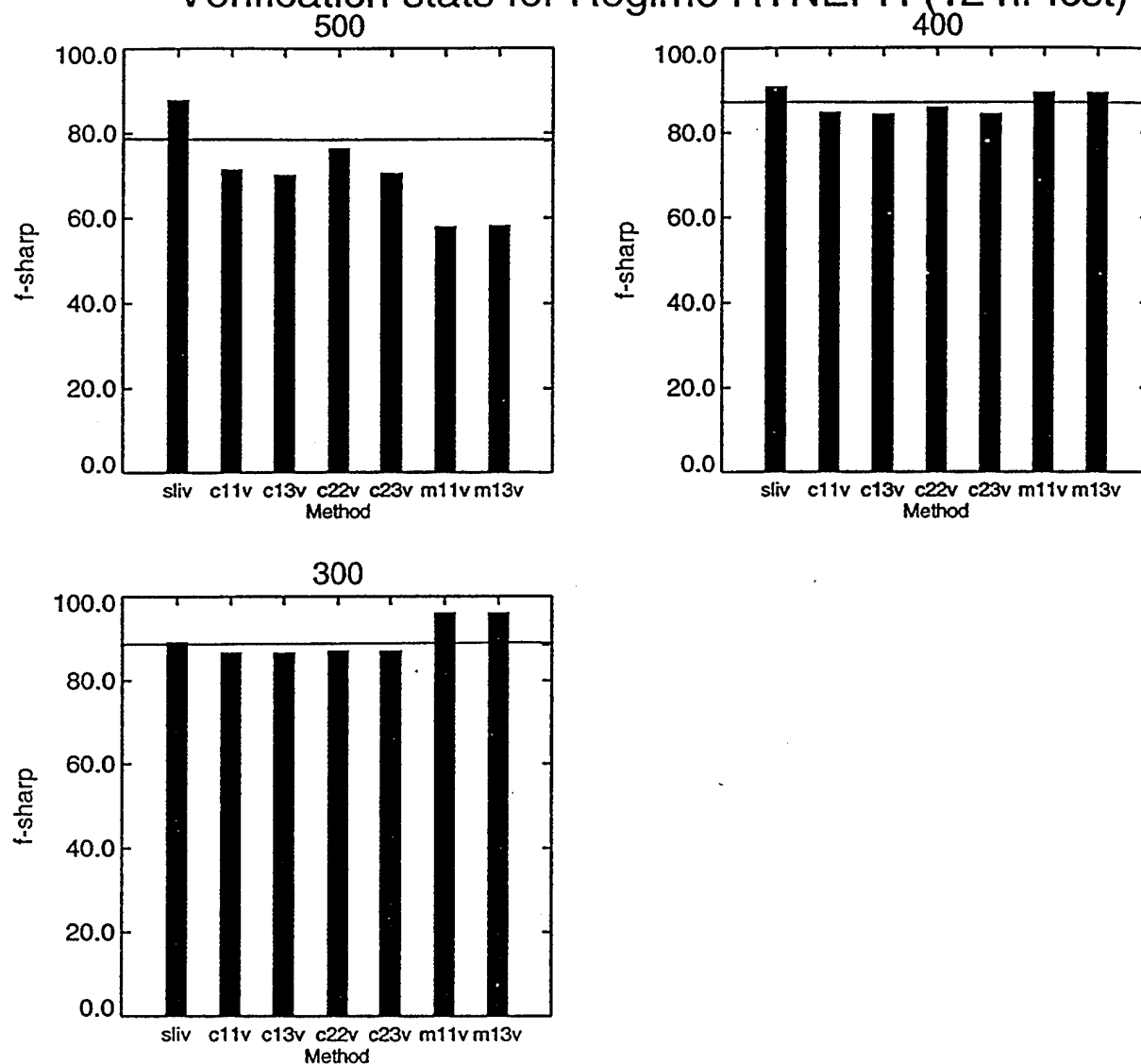


Figure 51: Continued.

# Verification stats for Regime RTNEPH (12 hr fcst)

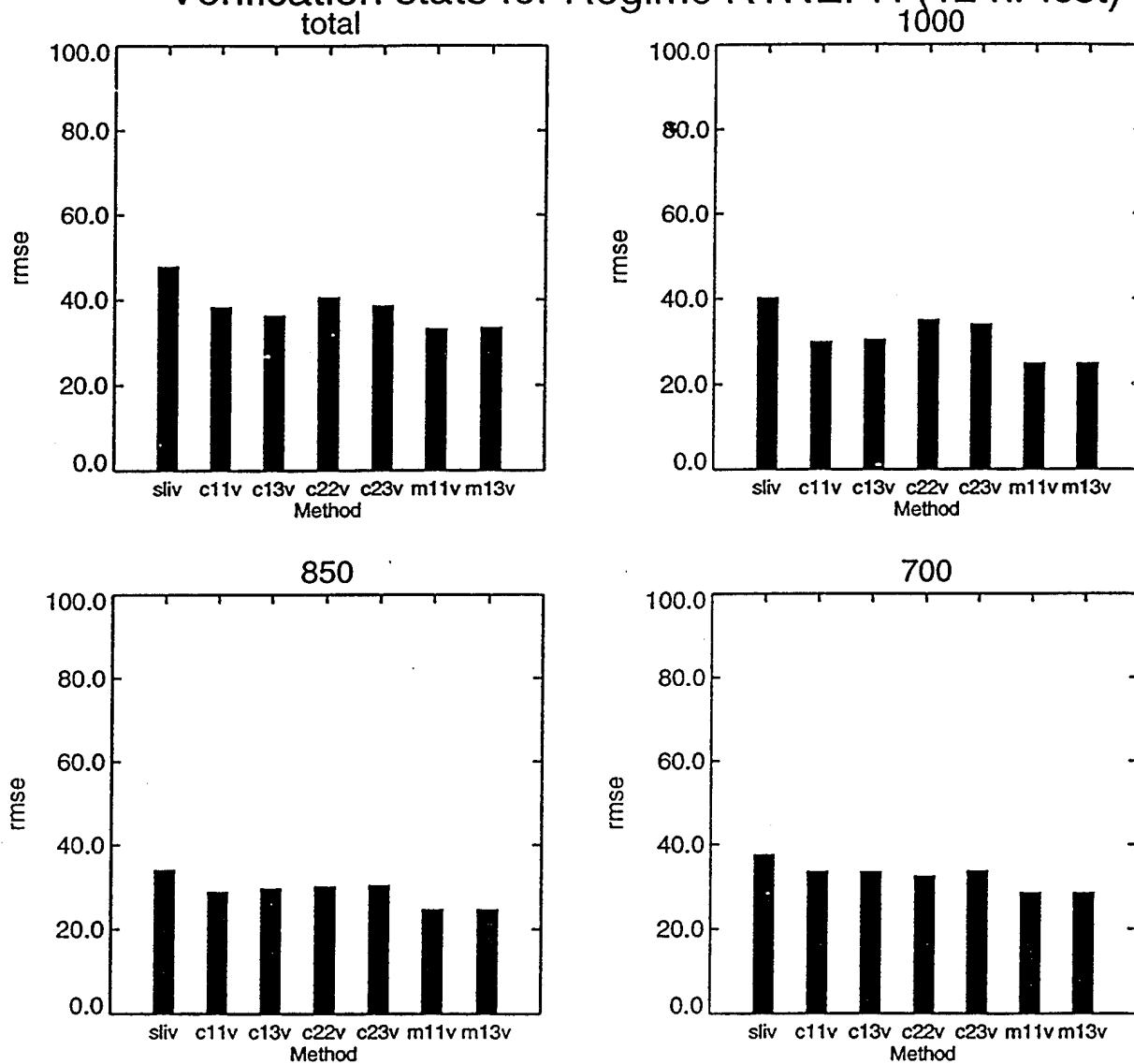


Figure 52: As Figure 50, except for the RMS error of 12-hour cloud cover forecasts.

# Verification stats for Regime RTNEPH (12 hr fcst)

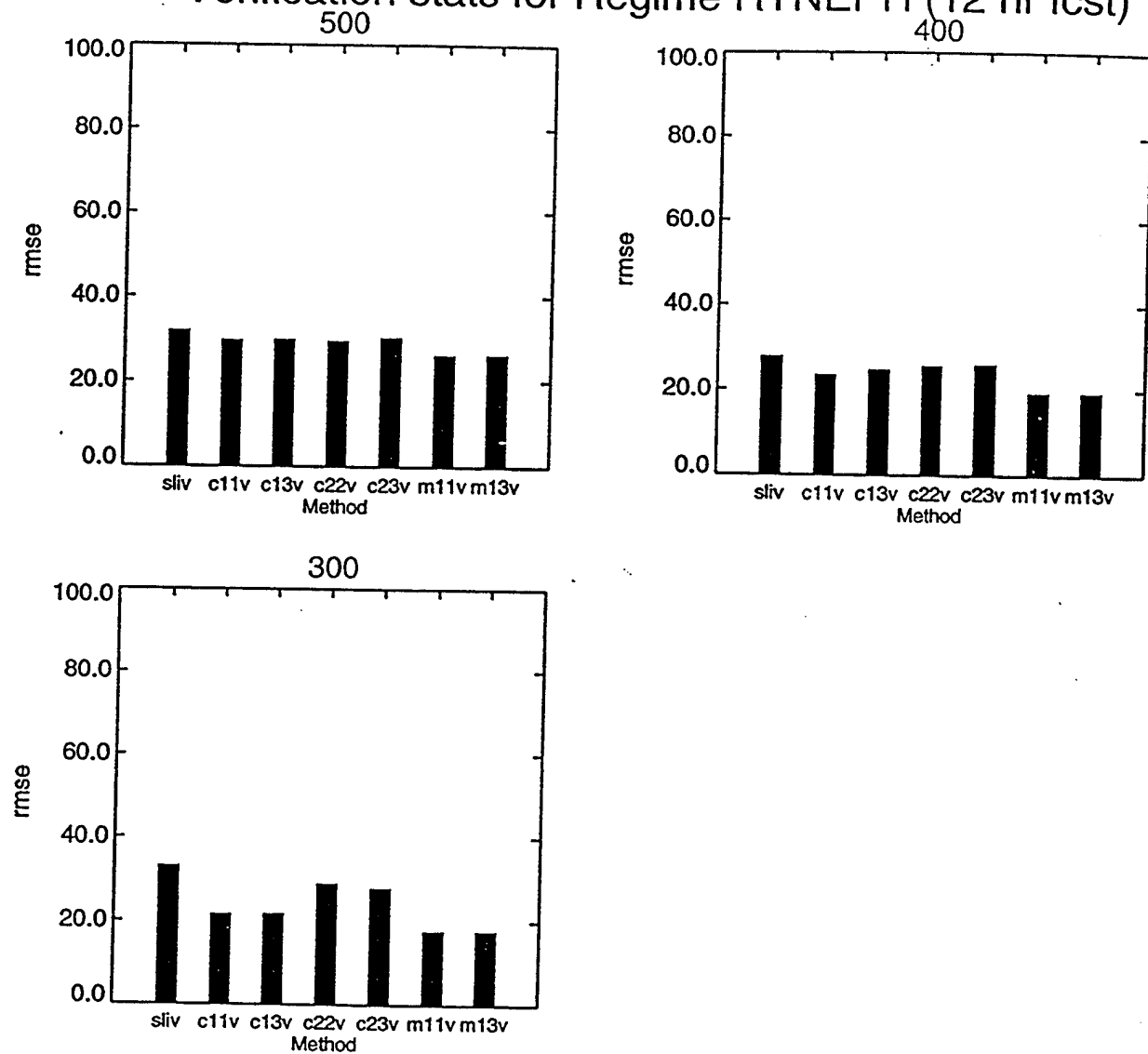


Figure 52: Continued.



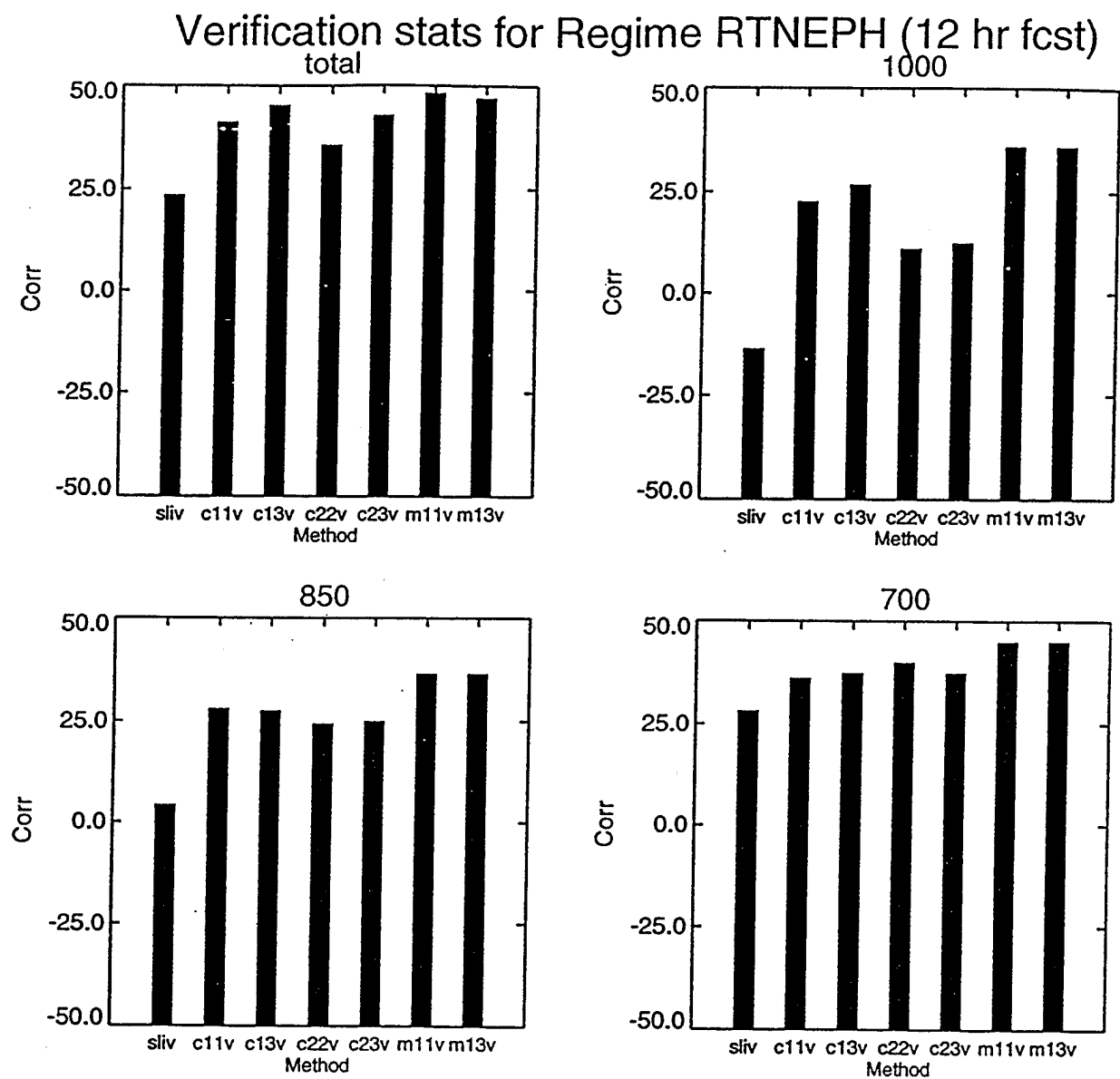


Figure 53: As Figure 50, except for the correlation between forecast (12 hour) and verifying cloud cover.

# Verification stats for Regime RTNEPH (12 hr fcst)

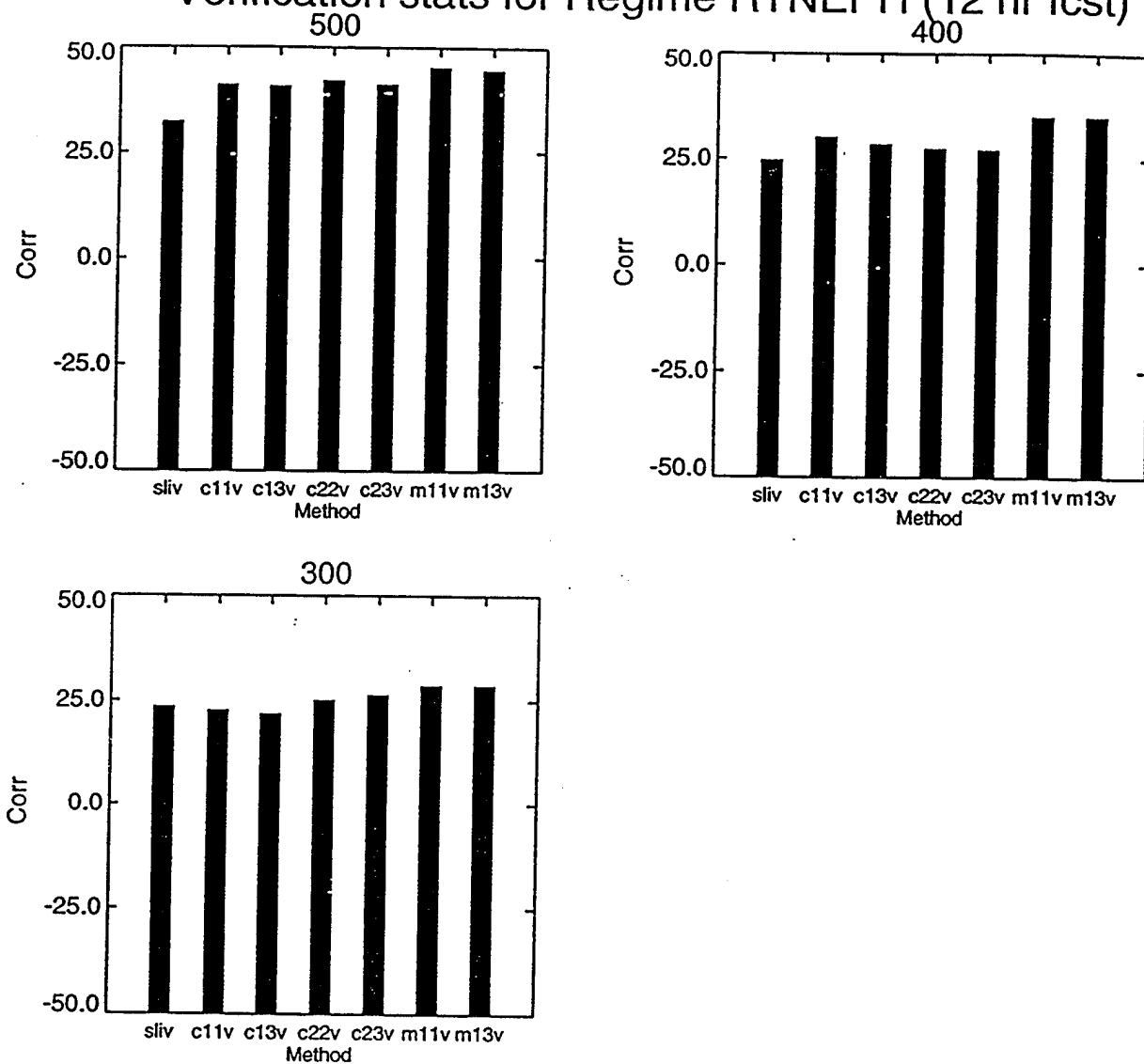


Figure 53: Continued.

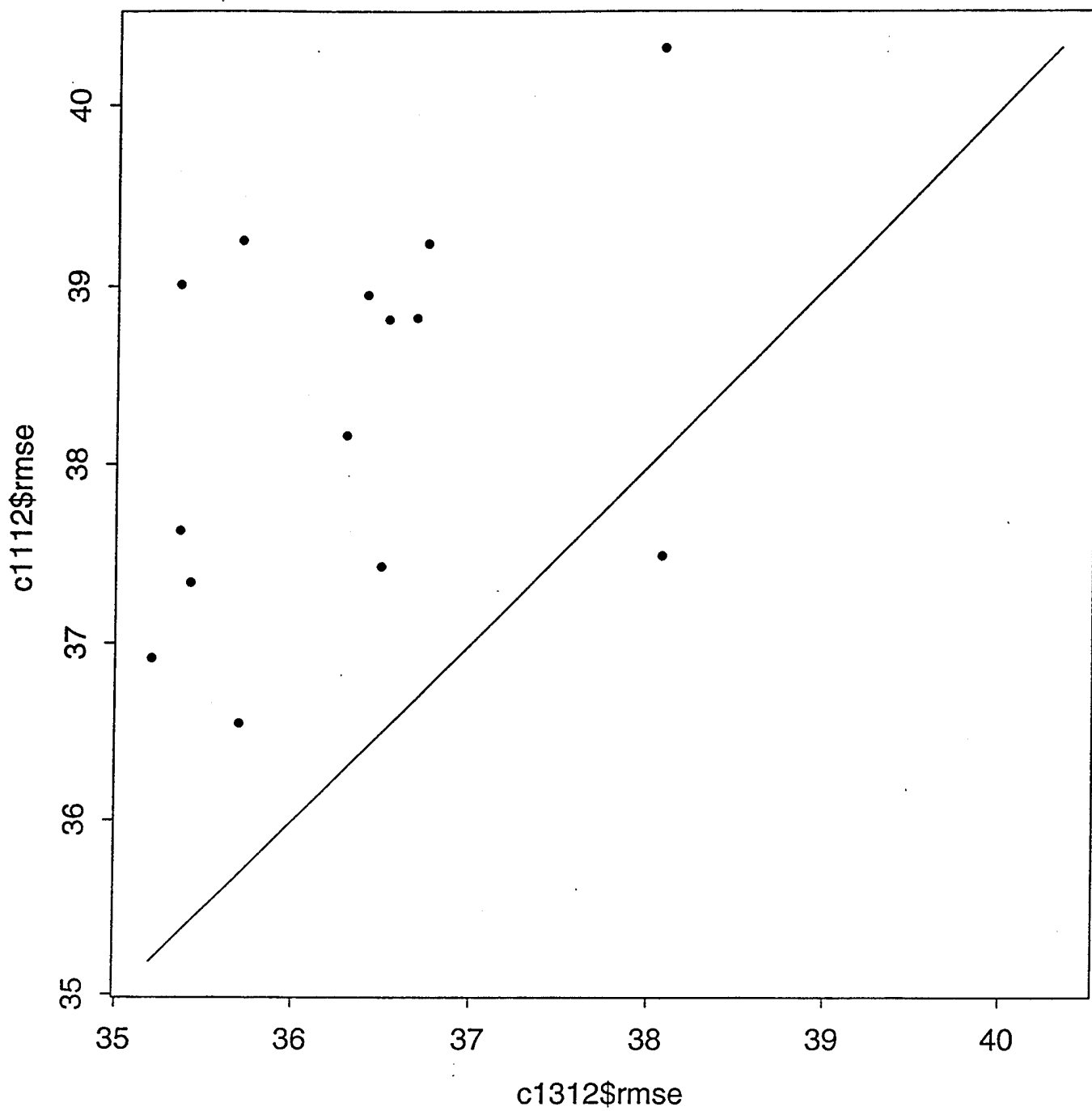
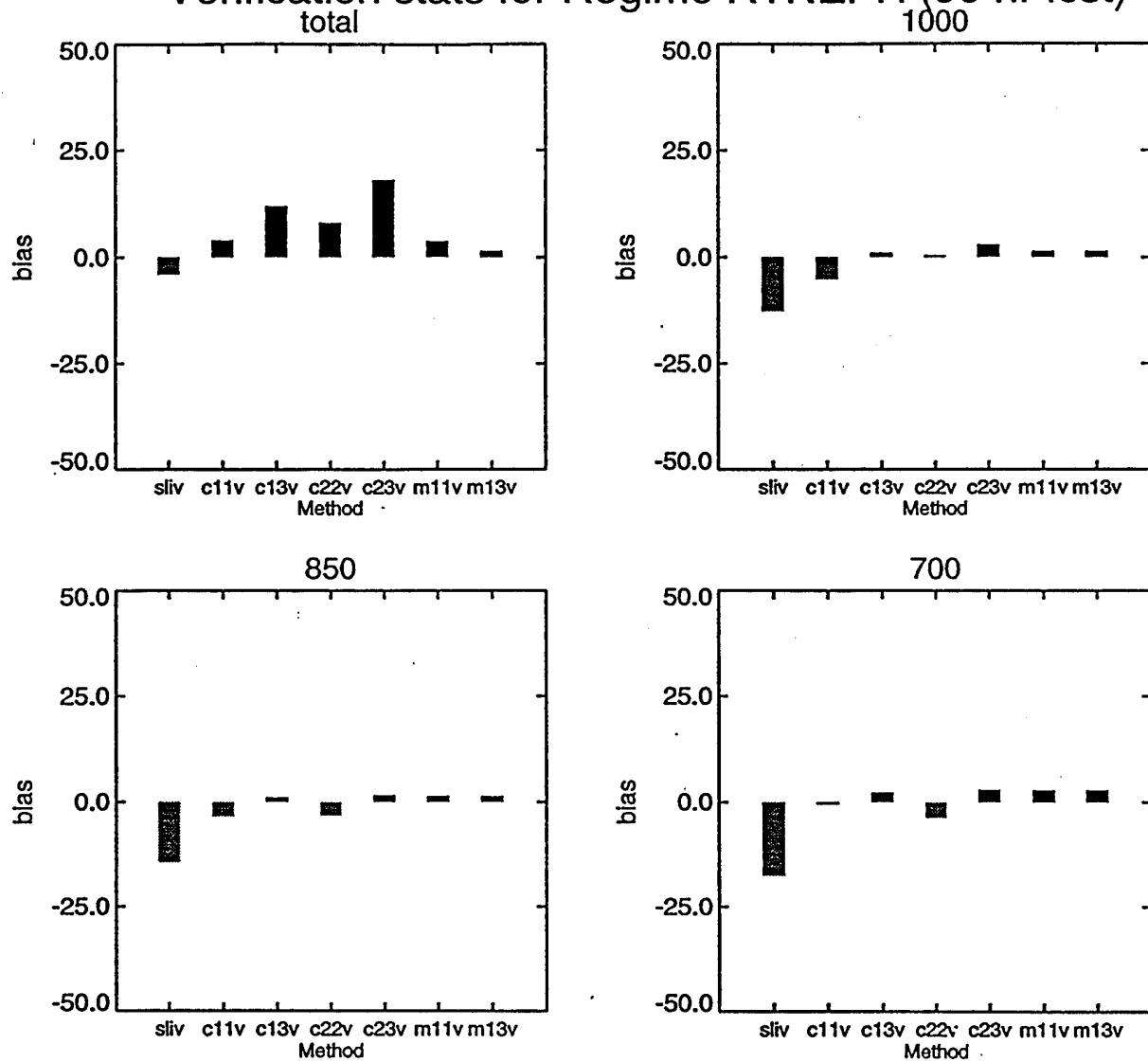


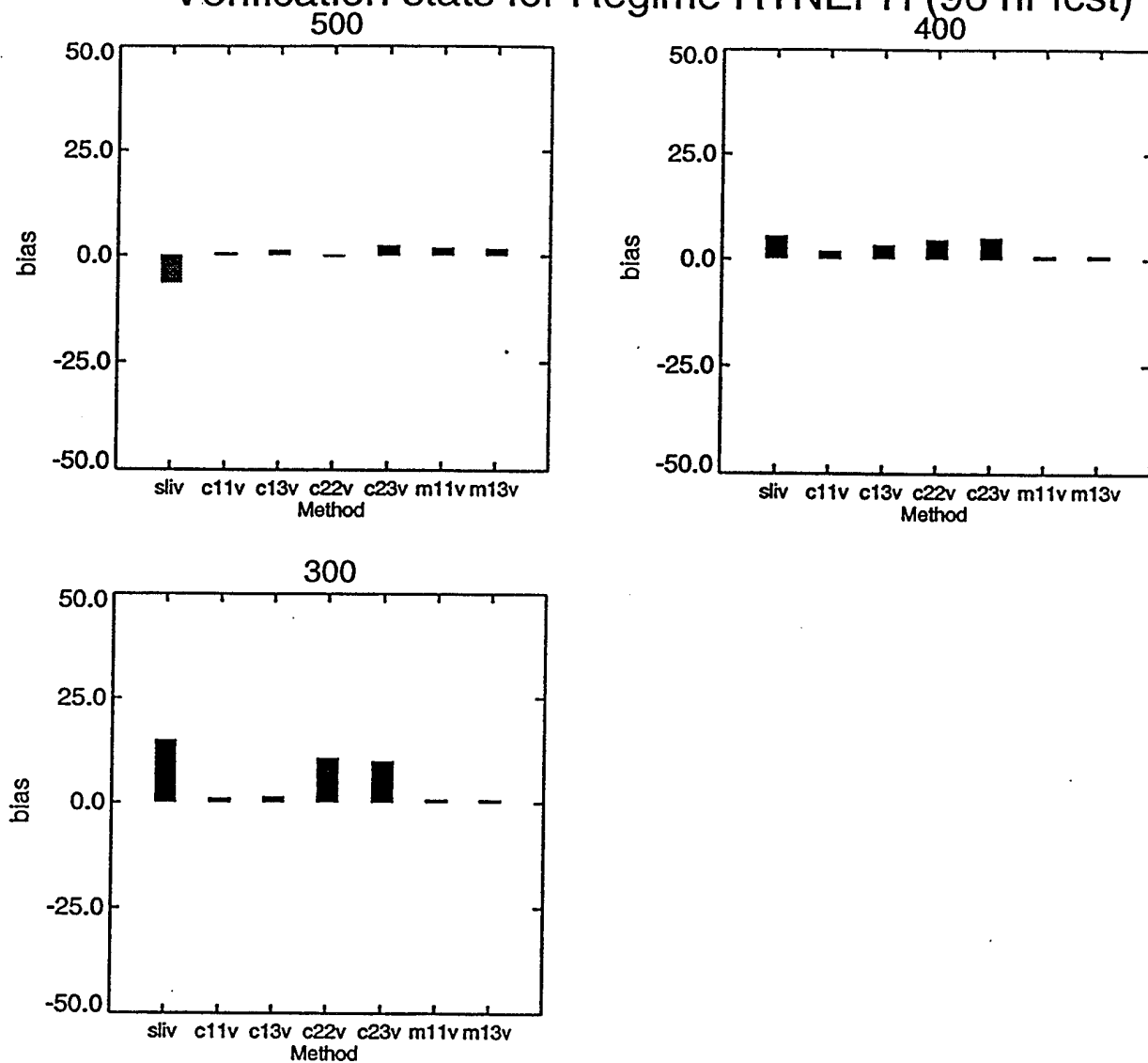
Figure 54: Scatterplot of 12-hour total cloud cover RMS errors of average RH CCA schemes. Scores for the scheme with separate total cloud cover CCA curves (c11) shown as function of the scores for the scheme with overlap computation (c13).

# Verification stats for Regime RTNEPH (96 hr fcst)



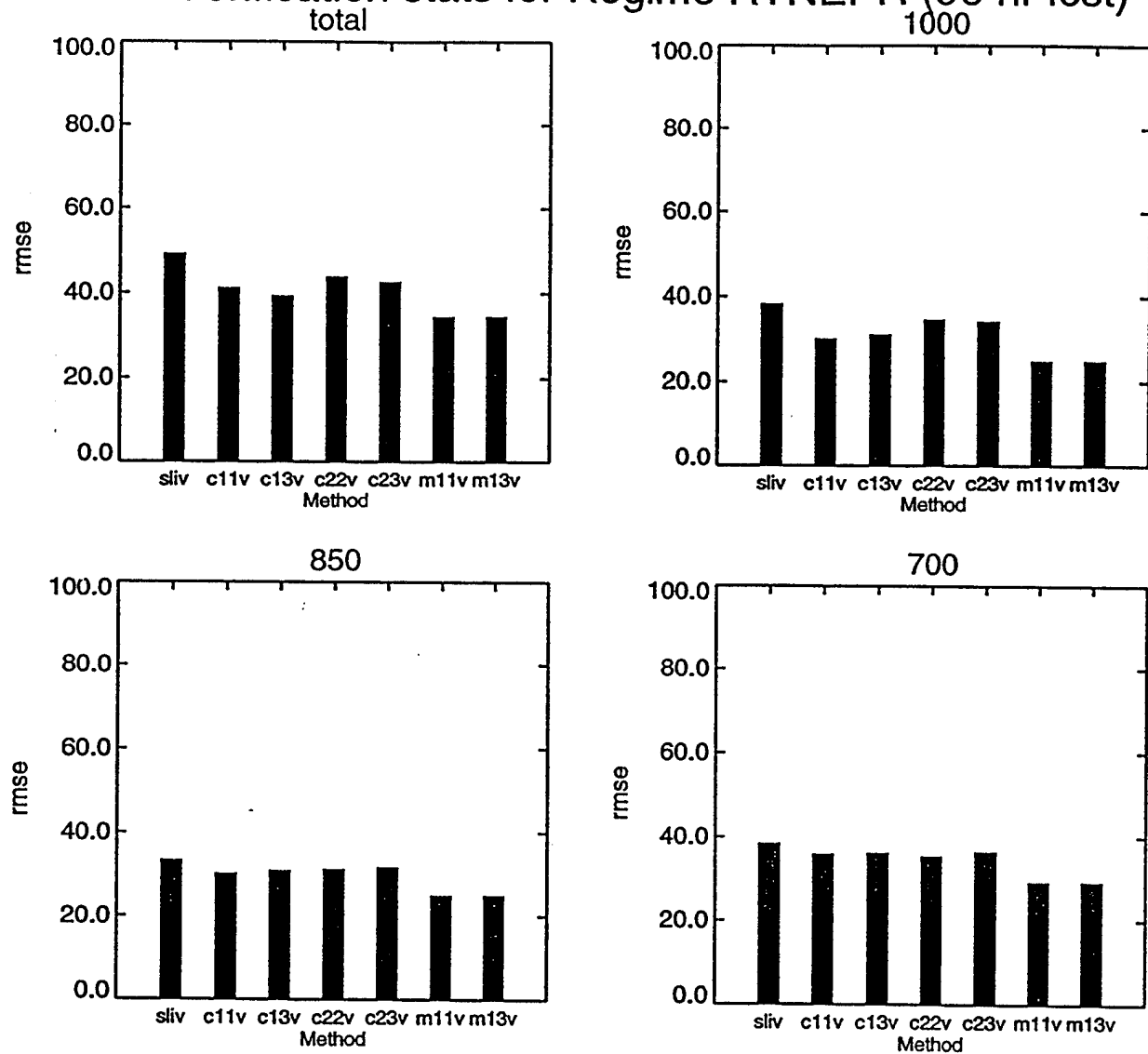
Figures 55: As Figure 50, except for the bias of 96-hour cloud cover forecasts.

# Verification stats for Regime RTNEPH (96 hr fcst)



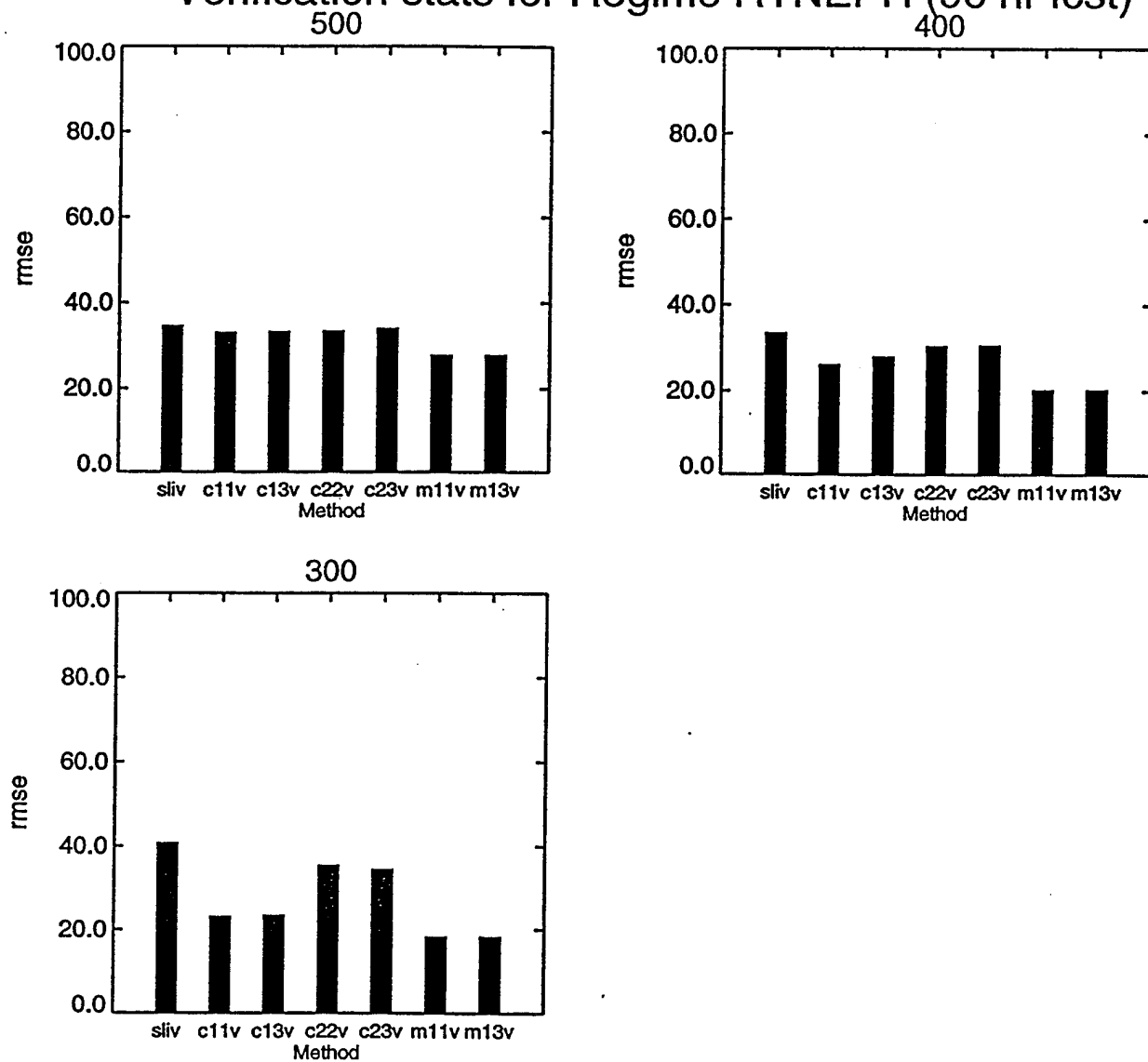
Figures 55: Continued.

# Verification stats for Regime RTNEPH (96 hr fcst)

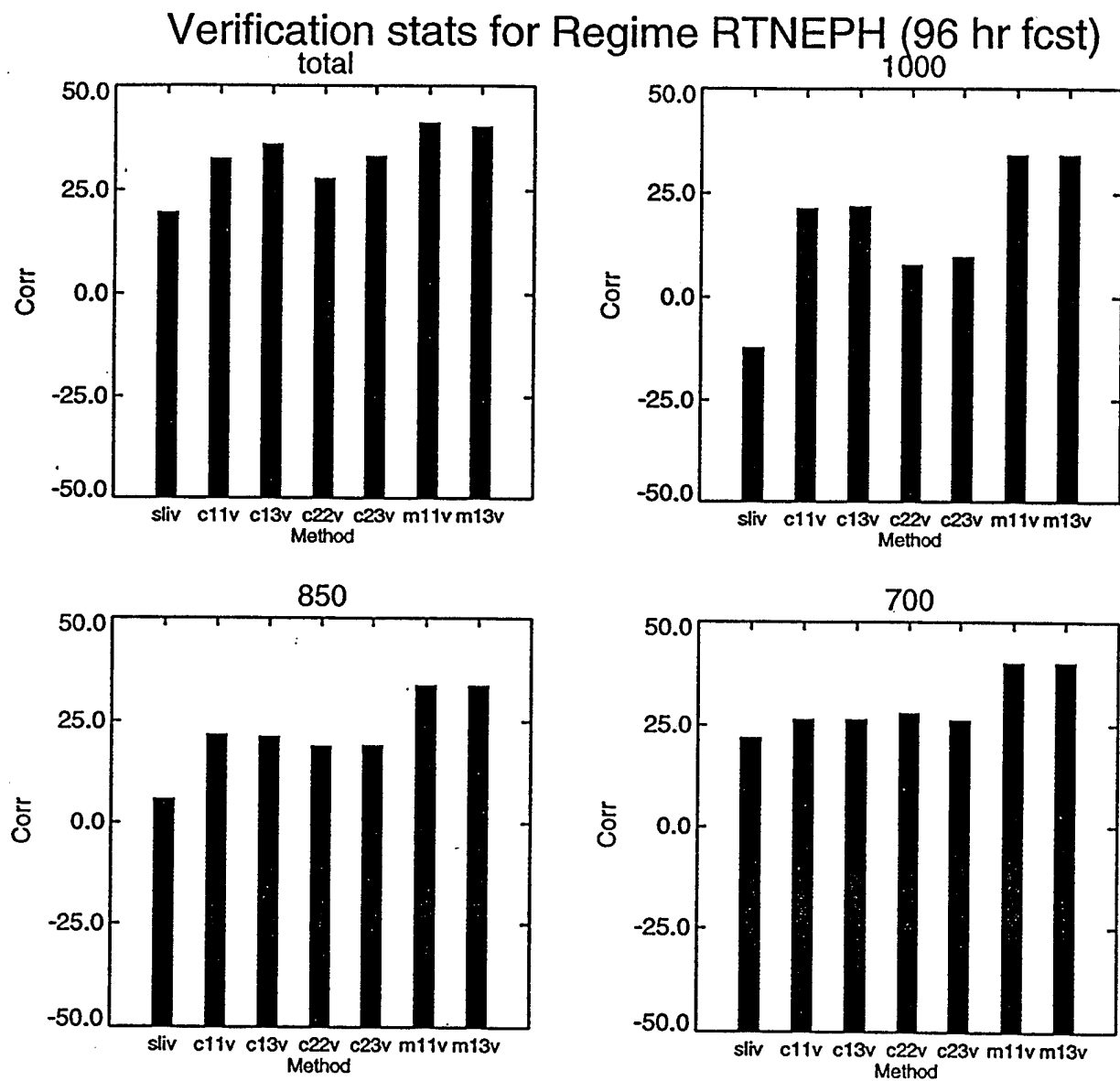


Figures 56: As Figure 50, except for the RMS errors of 96-hour cloud cover forecasts.

# Verification stats for Regime RTNEPH (96 hr fcst)



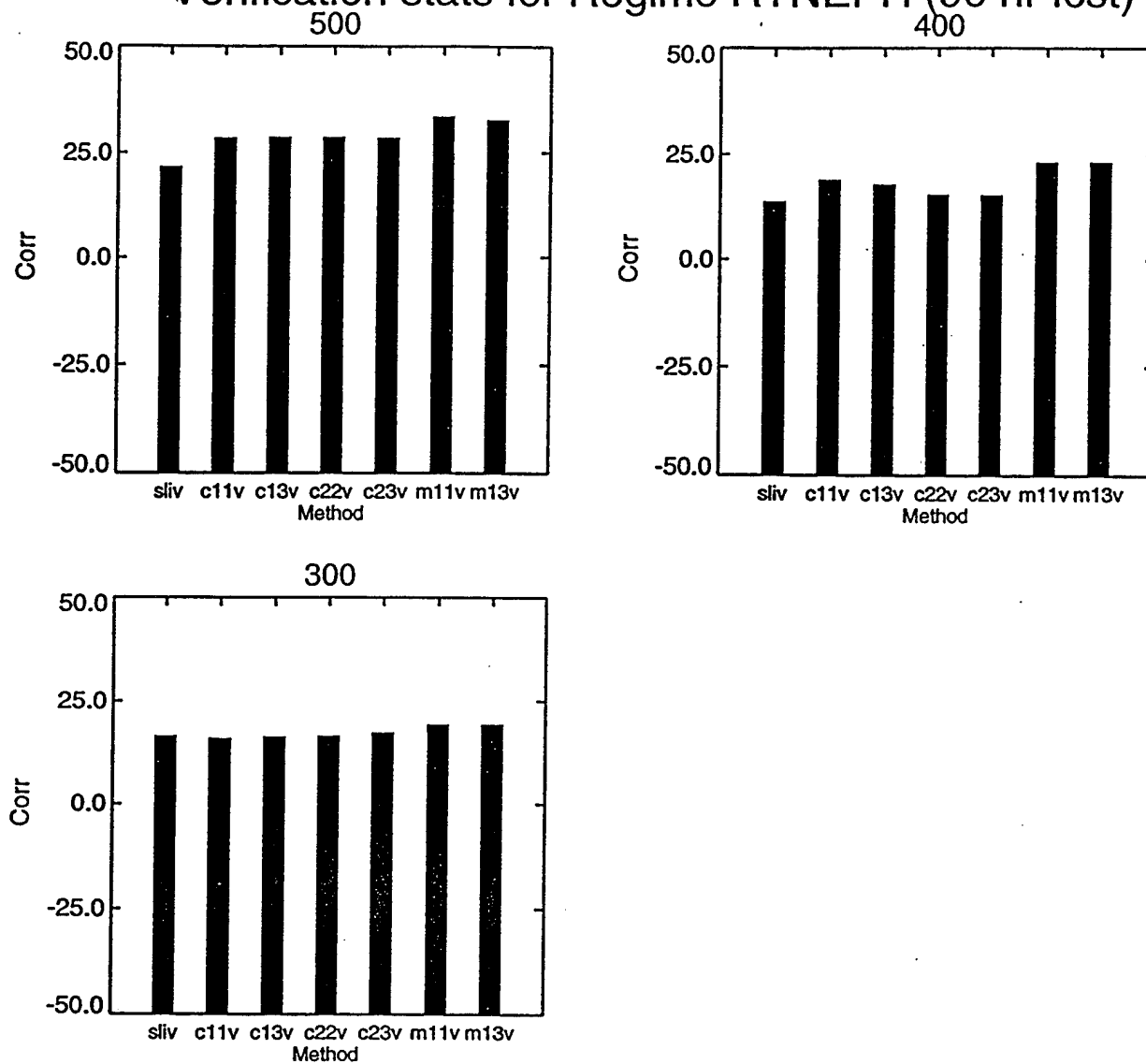
Figures 56: Continued.



Figures 57: As Figure 50, except for the correlation between forecast (96-hour) and verifying cloud cover.



# Verification stats for Regime RTNEPH (96 hr fcst)



Figures 57: Continued.

# January Verification stats for Regime RTNEPH

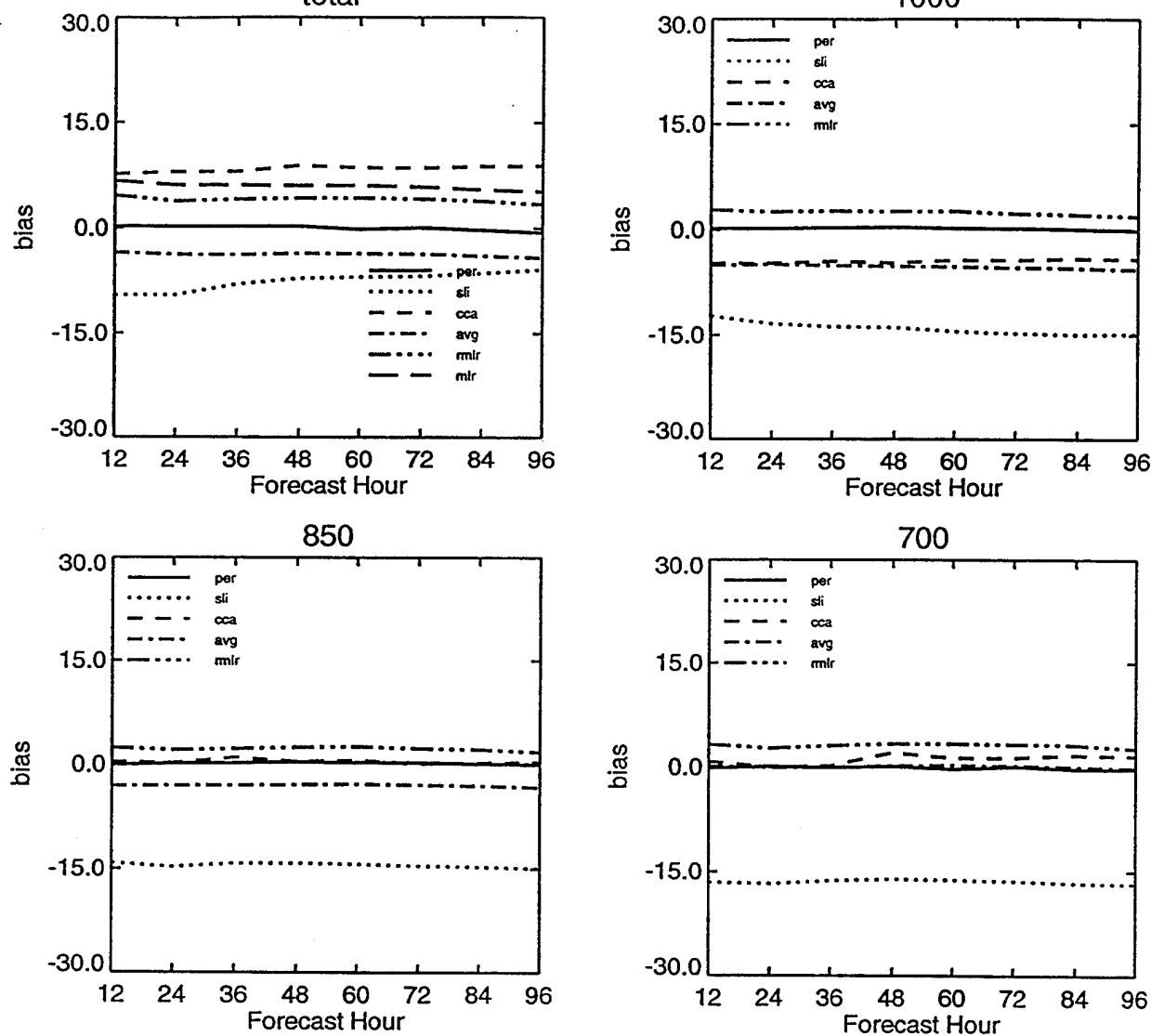


Figure 58: Bias of 12-96 hour forecasts of cloud cover (Total and the six RTNEPH layers) of the 6 cloud cover schemes for January, computed over the NH RTNEPH octagon. Refer to text for explanation of curve labels.

# January Verification stats for Regime RTNEPH

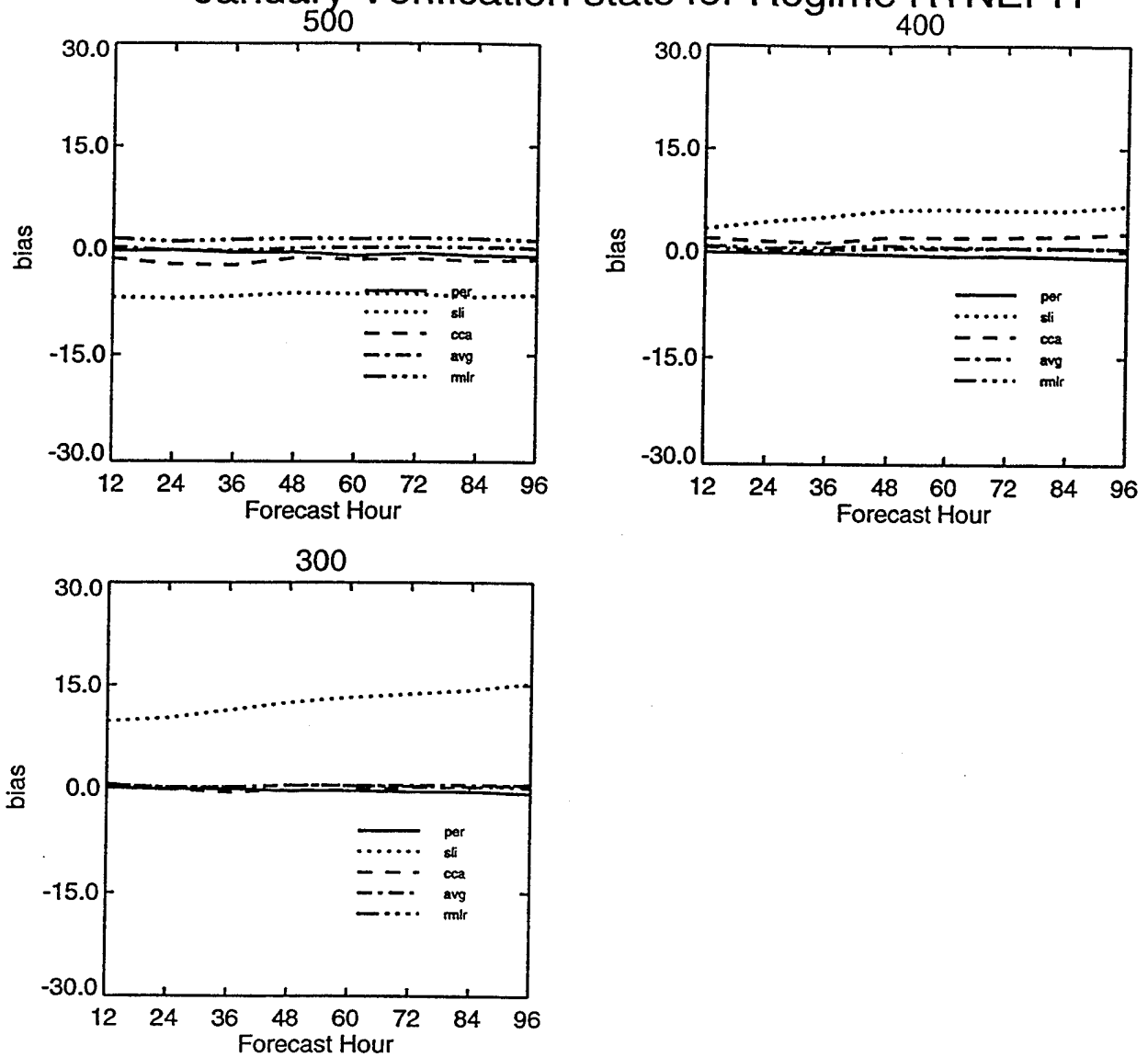


Figure 58: Continued.

# January Verification stats for Regime Fronthalf

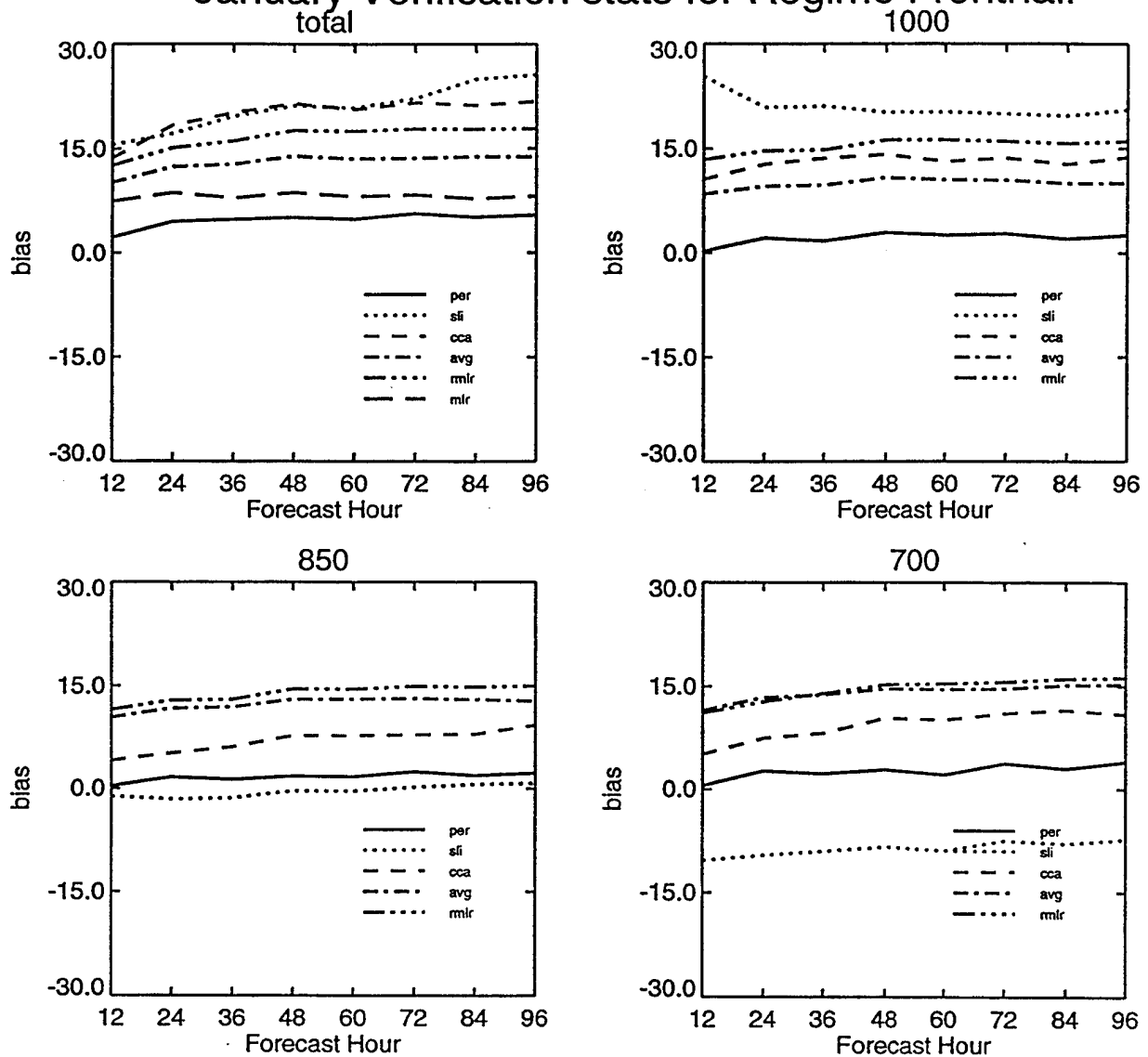


Figure 59: Same as Figure 58, except for the bias of 12-96 hour forecasts of cloud cover, computed over the "fronthalf" portion of the NH RTNEPH octagon.

# January Verification stats for Regime Fronthalf

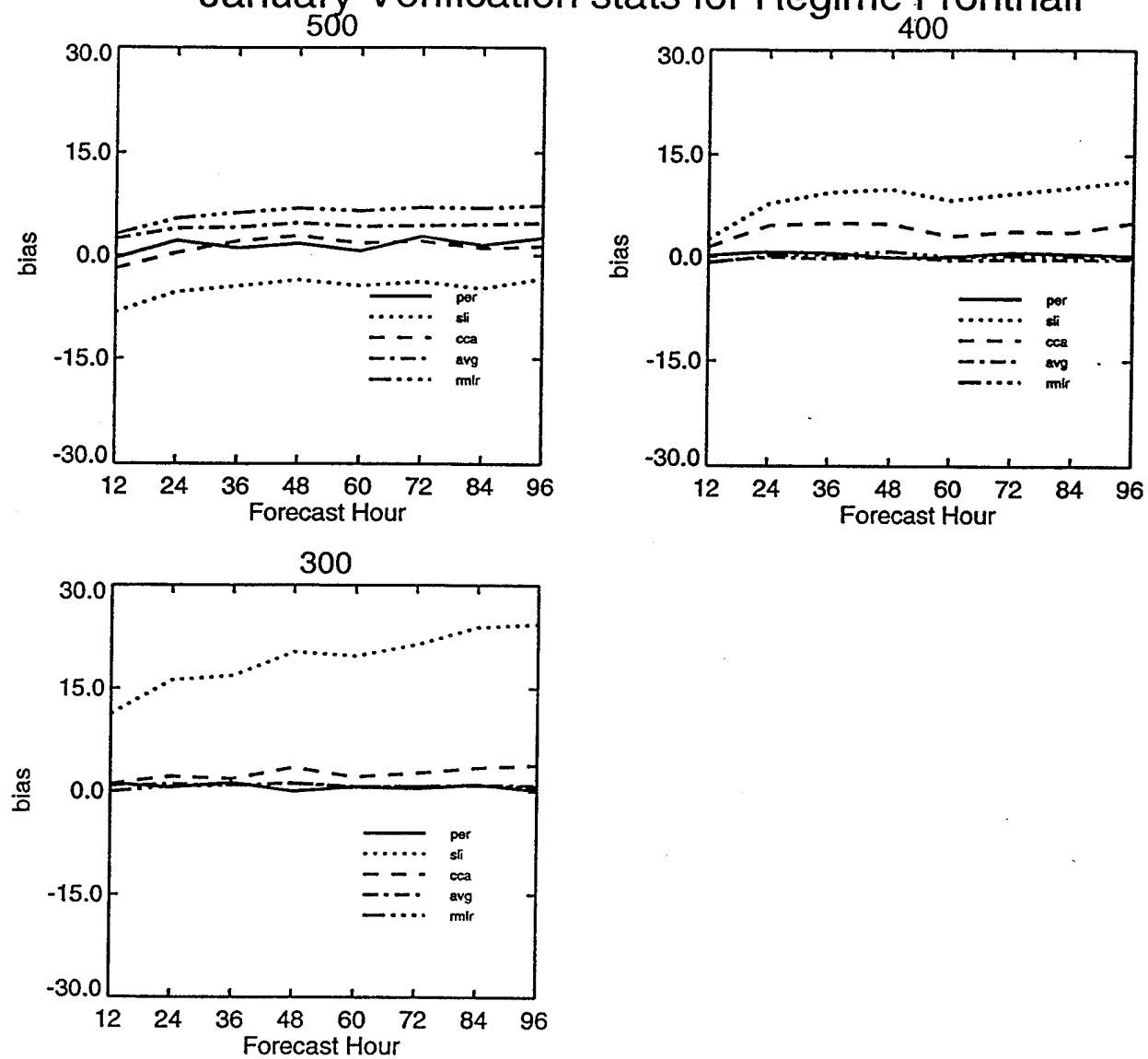


Figure 59: Continued.

# January Verification stats for Regime Backhalf

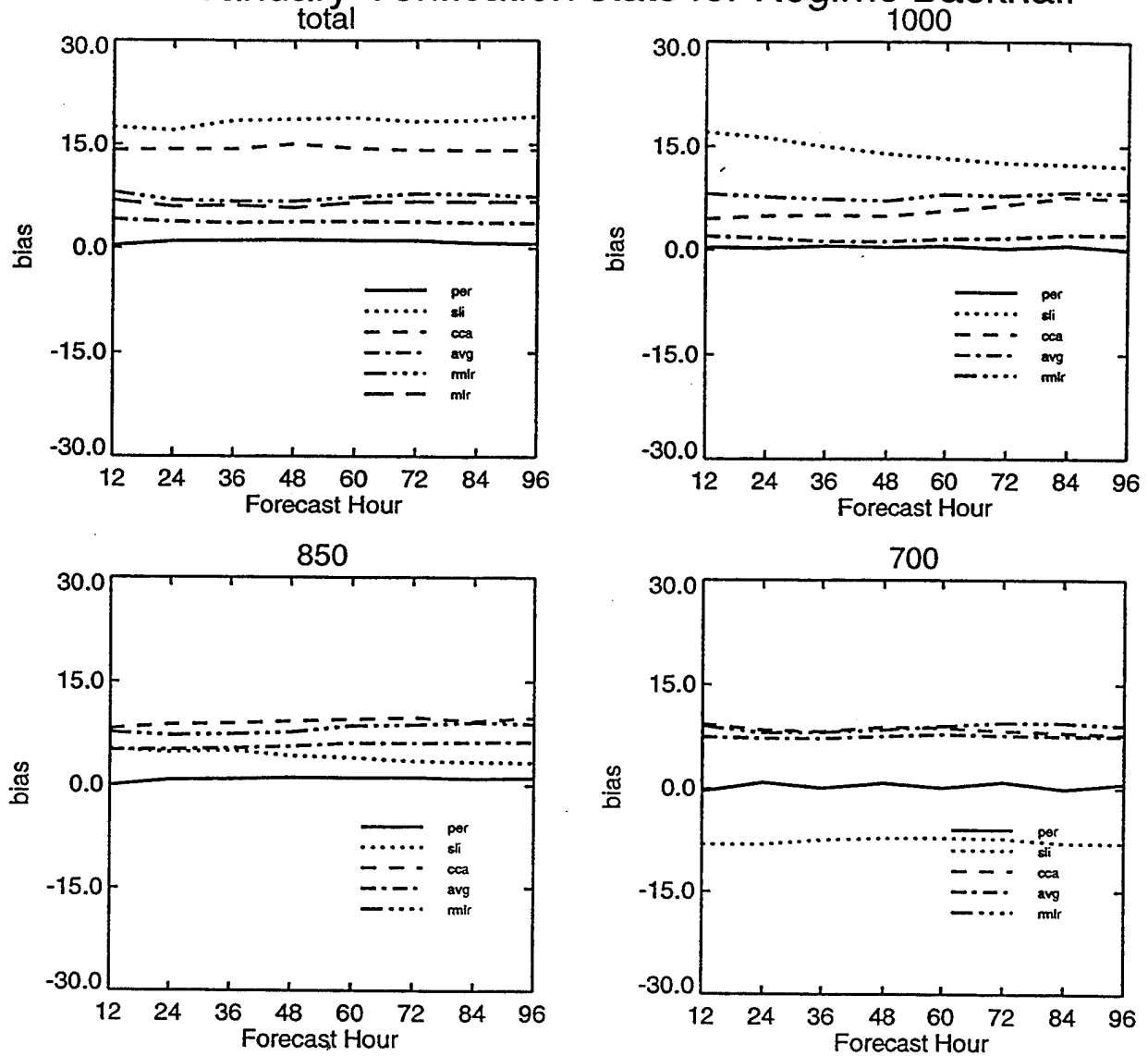


Figure 60: Same as Figure 58, except for the bias of 12-96 hour forecasts of cloud cover, computed over the "backhalf" portion of the NH RTNEPH octagon.

# January Verification stats for Regime Backhalf

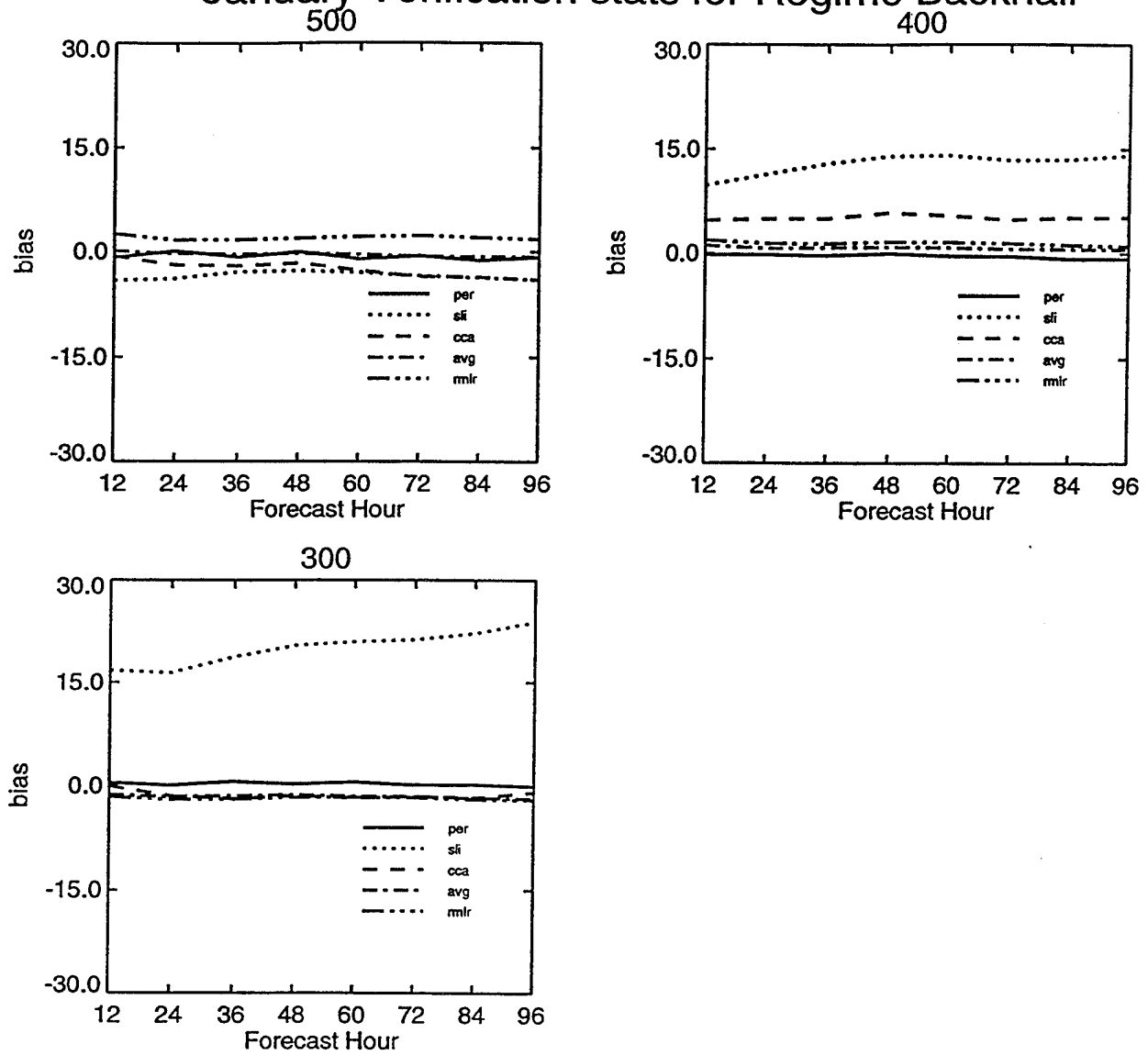


Figure 60: Continued.

# January Verification stats for Regime RTNEPH

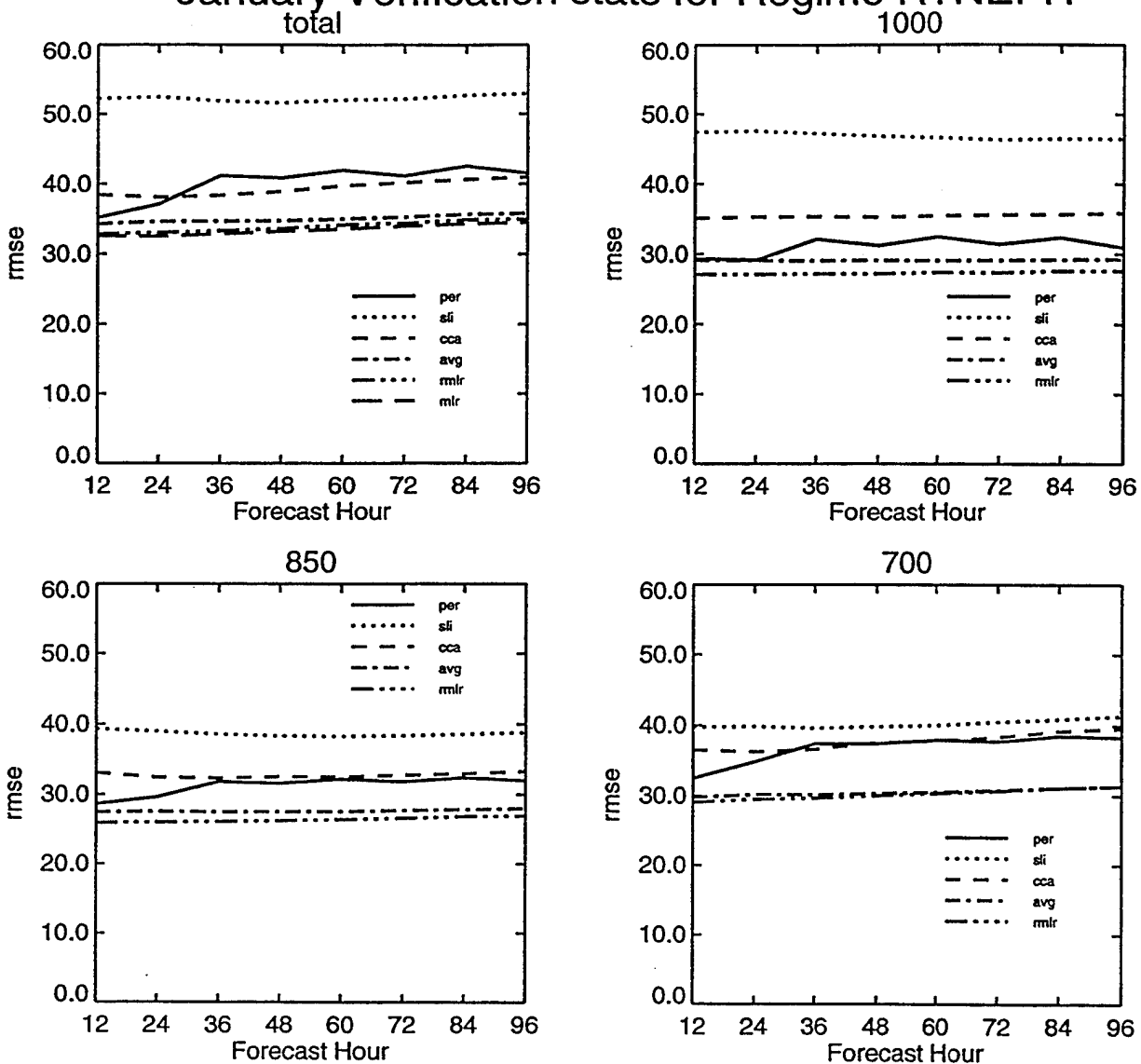


Figure 61: Same as Figure 58, except for the rmse of 12-96 hour forecasts of cloud cover, computed over the NH RTNEPH octagon.



# January Verification stats for Regime RTNEPH

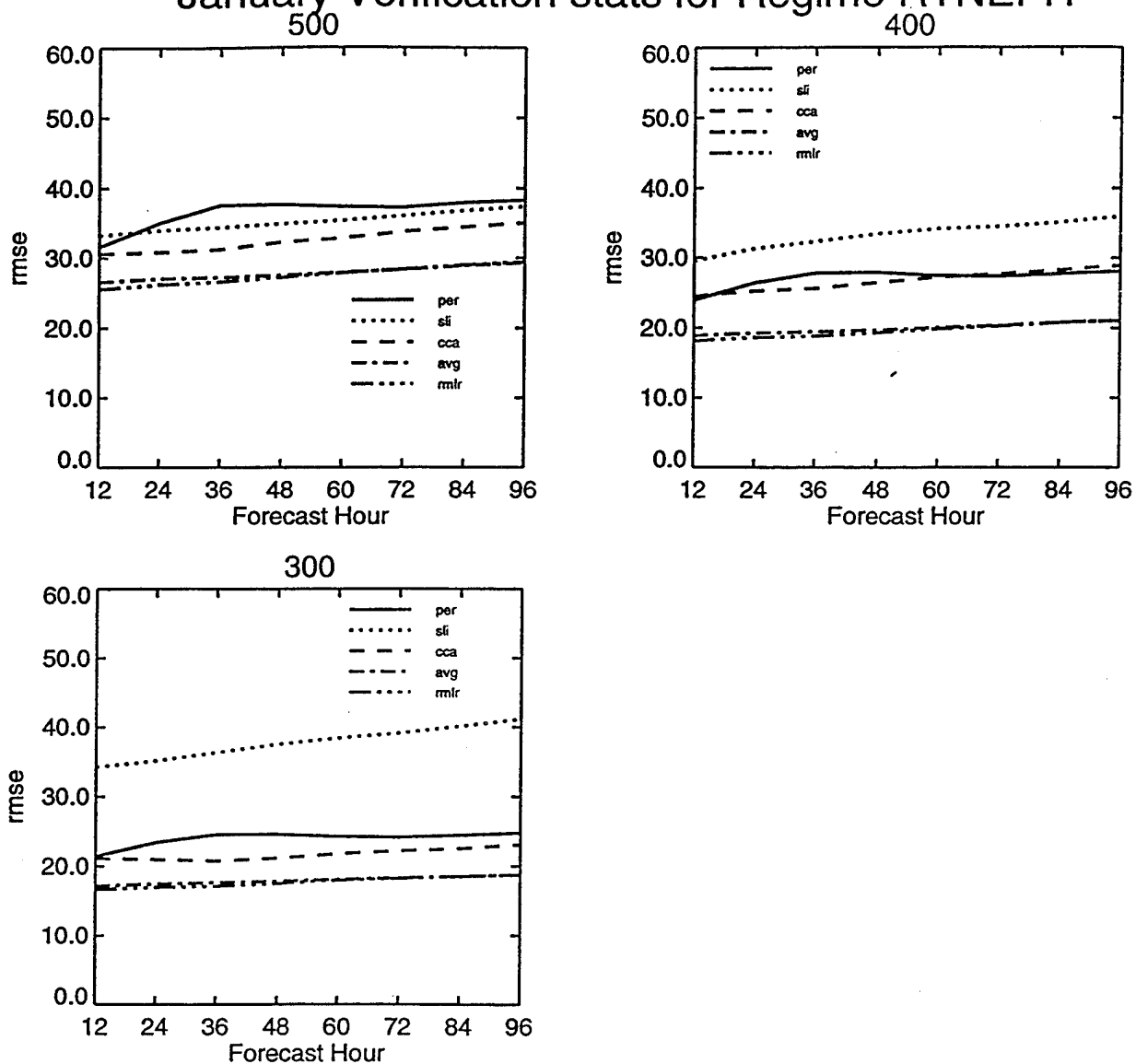


Figure 61: Continued.

## January Verification stats for Regime Fronthalf

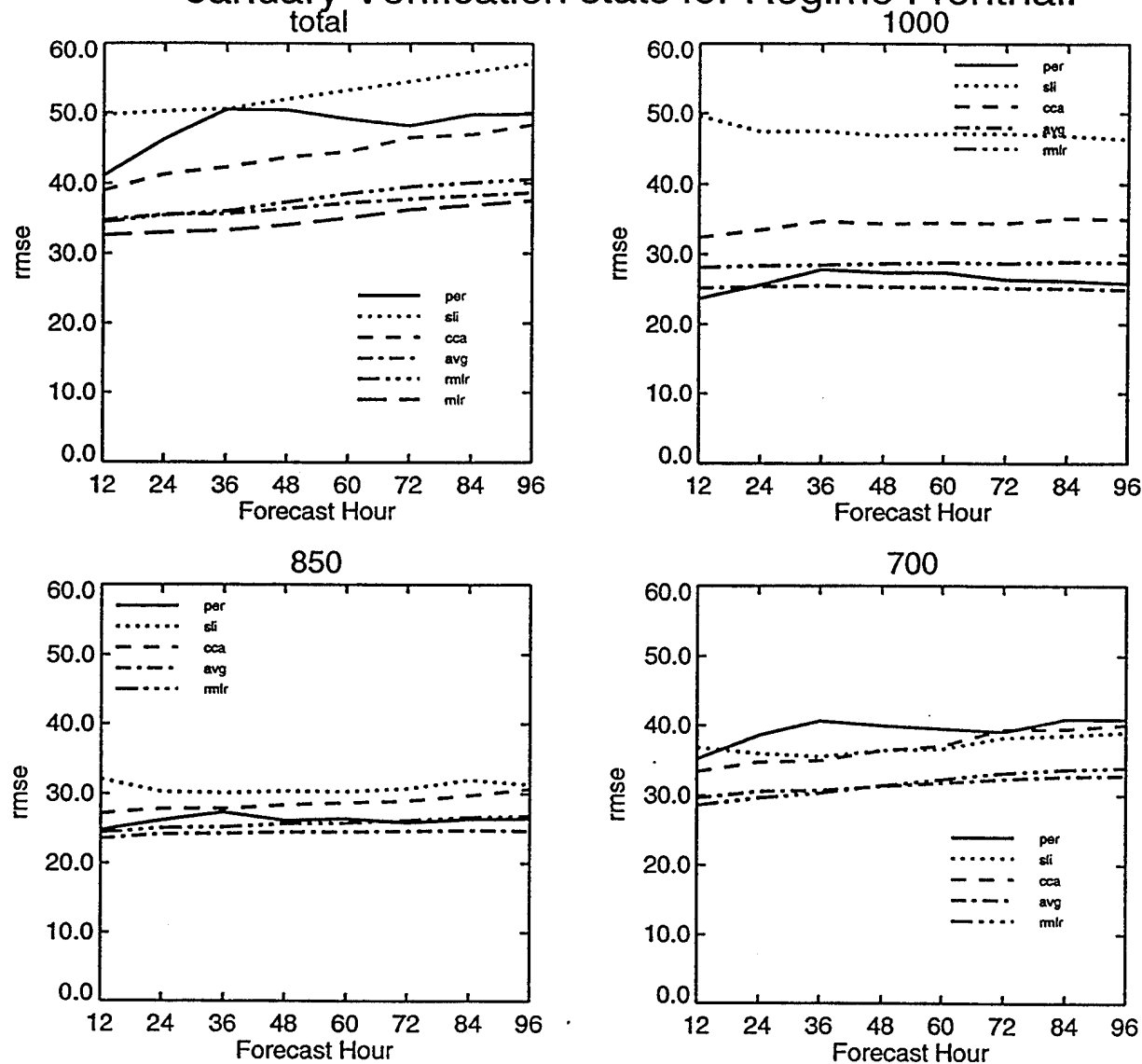


Figure 62: Same as Figure 58, except for the rmse of 12-96 hour forecasts of cloud cover, computed over the "fronthalf" portion of the NH RTNEPH octagon.

# January Verification stats for Regime Fronthalf

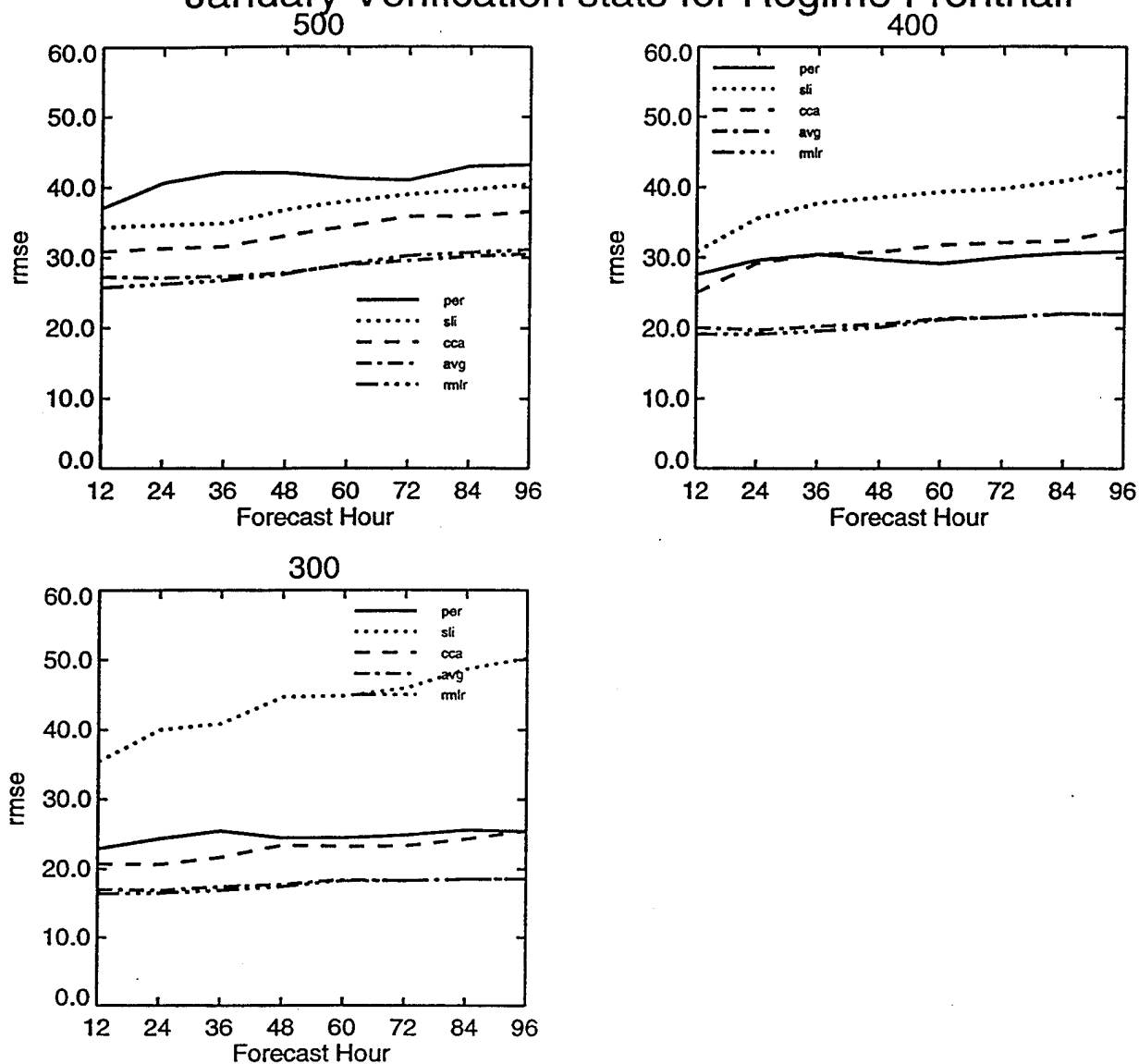


Figure 62: Continued.

# January Verification stats for Regime Backhalf

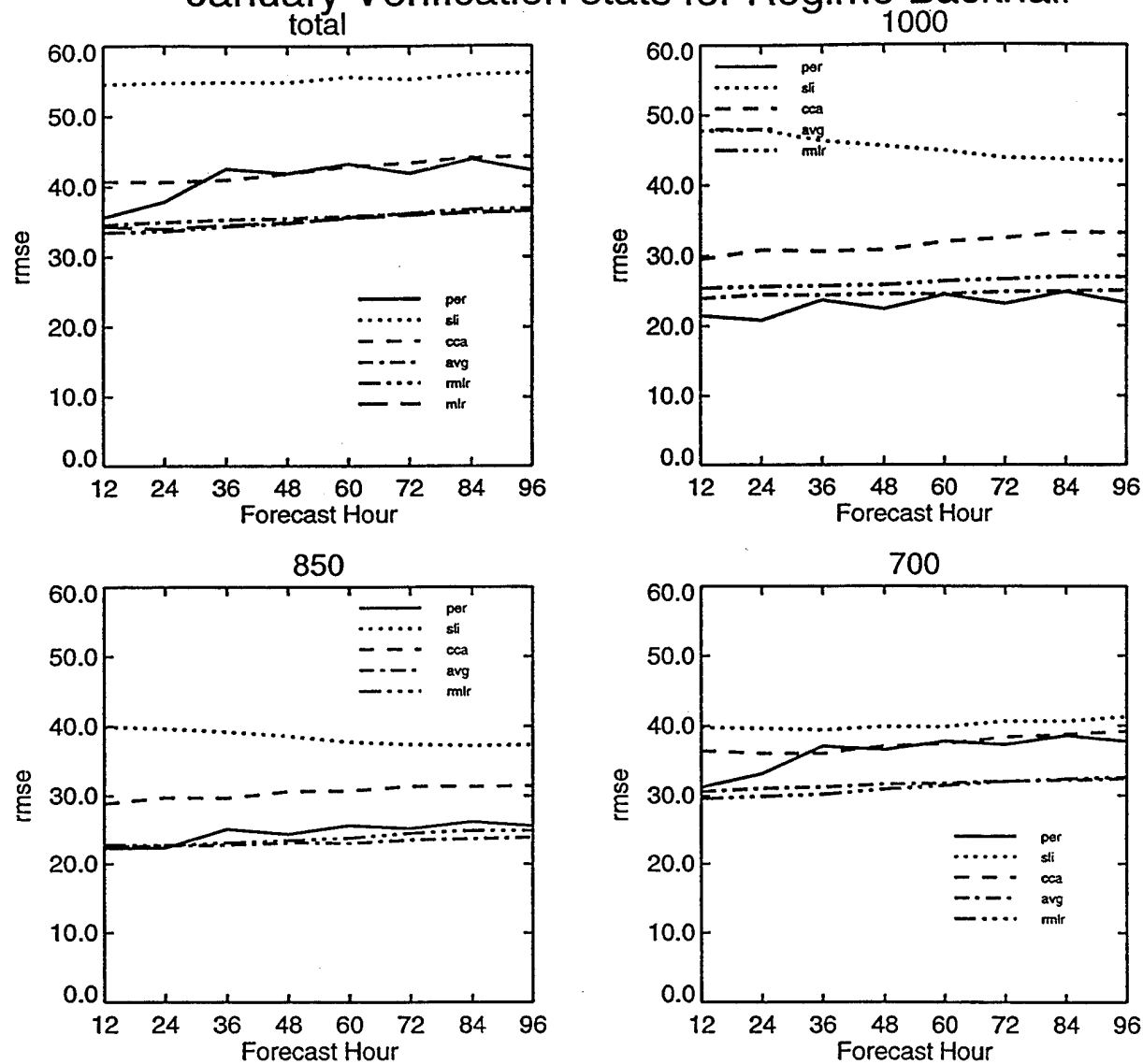


Figure 63: Same as Figure 58, except for the rmse of 12-96 hour forecasts of cloud cover, computed over the "backhalf" portion of the NH RTNEPH octagon.

# January Verification stats for Regime Backhalf

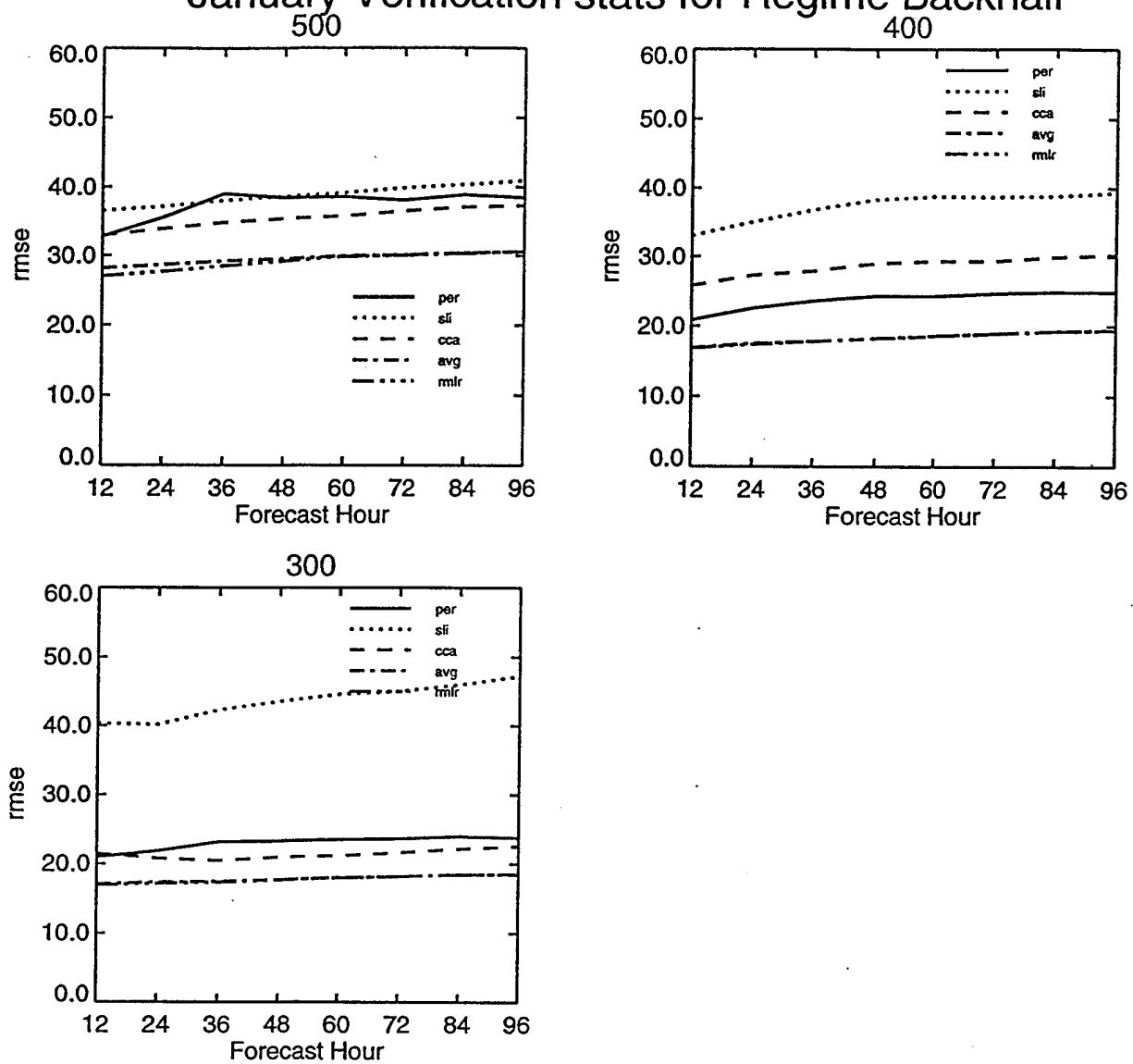


Figure 63: Continued.

# January Verification stats for Regime RTNEPH

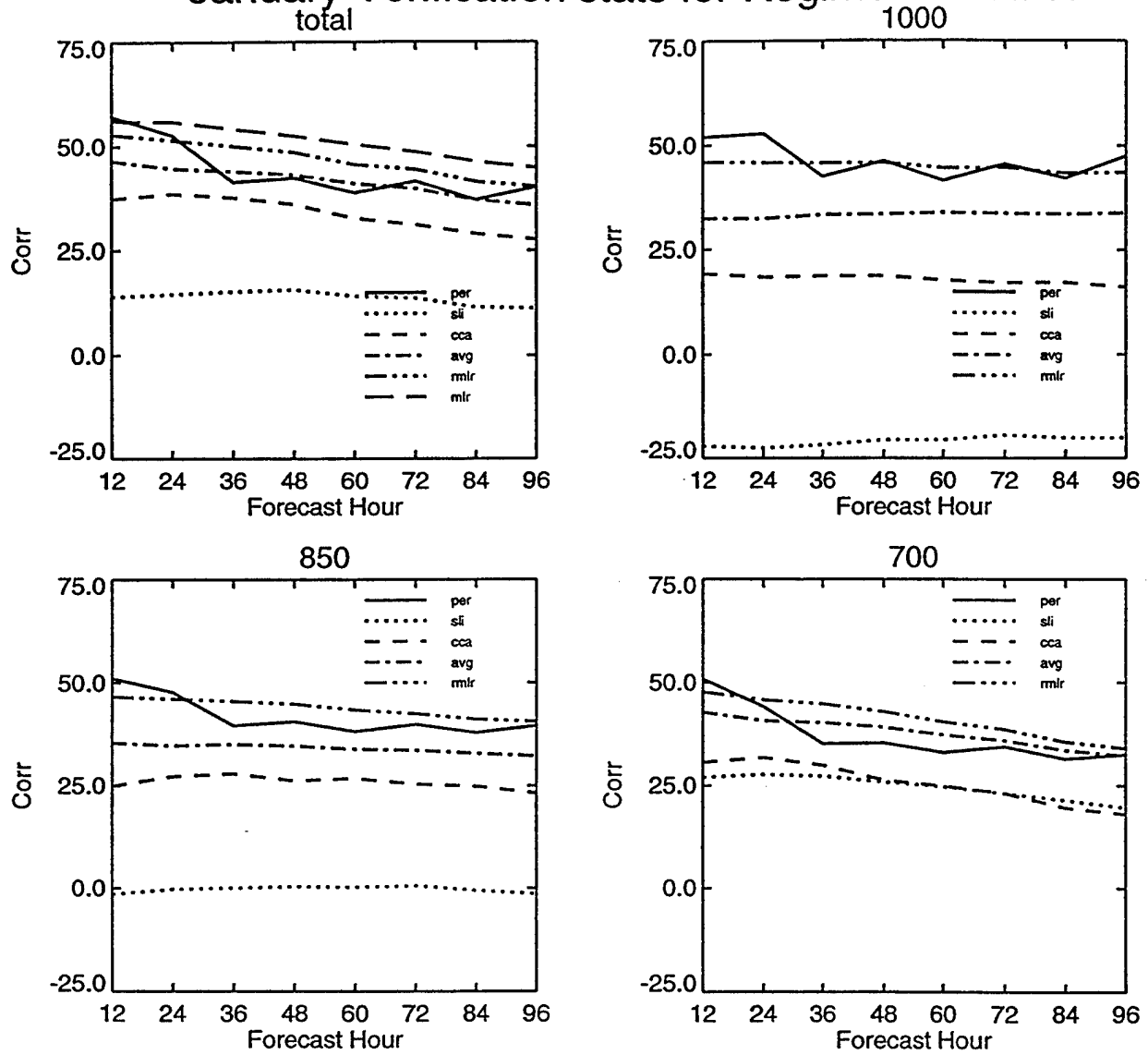


Figure 64: Same as Figure 58, except for the correlation of 12-96 hour forecasts of cloud cover, computed over the NH RTNEPH octagon.

# January Verification stats for Regime RTNEPH

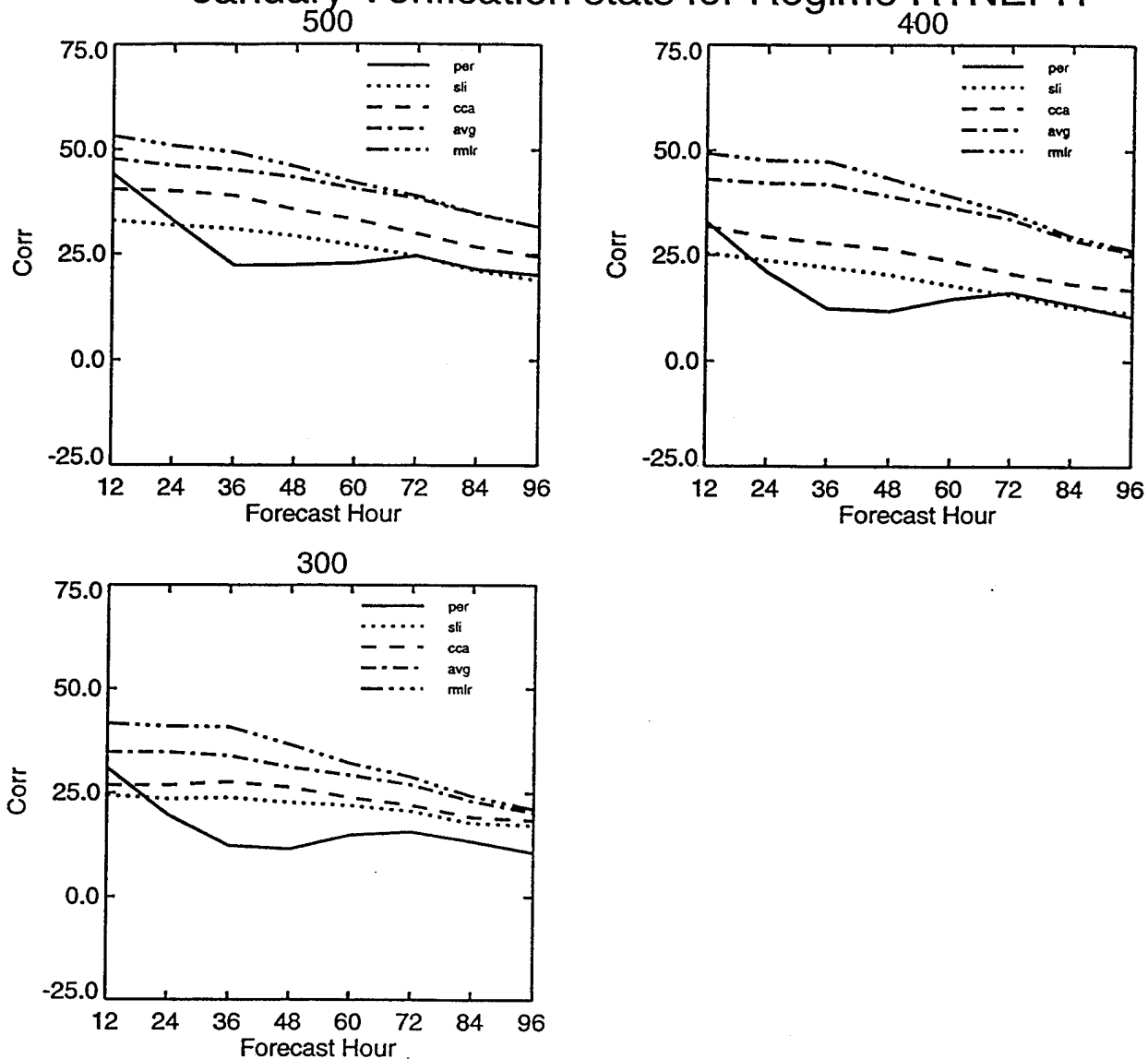


Figure 64: Continued.

# January Verification stats for Regime Fronthalf

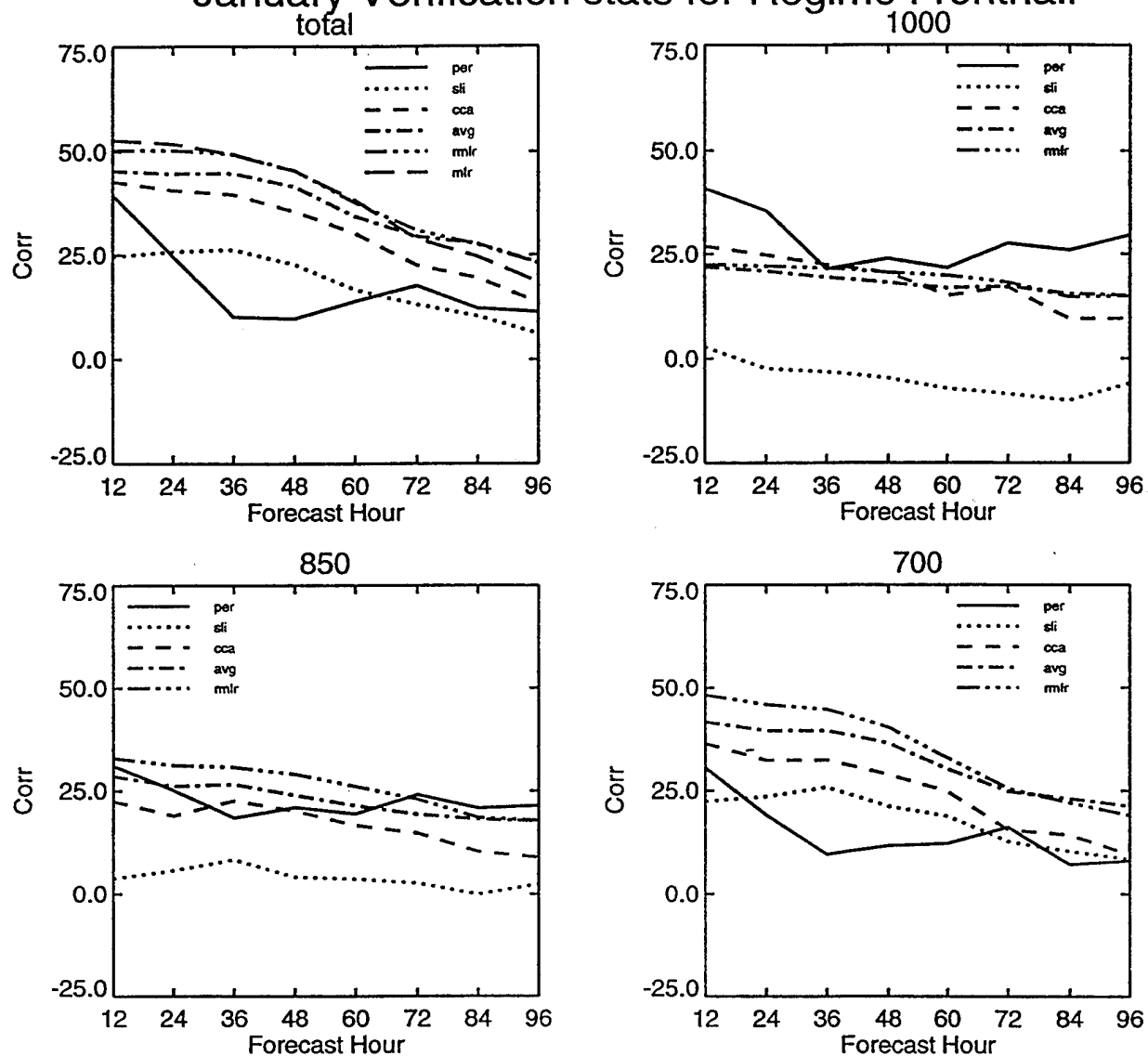


Figure 65: Same as Figure 58, except for the correlation of 12-96 hour forecasts of cloud cover, computed over the "fronthalf" portion of the NH RTNEPH octagon.



# January Verification stats for Regime Fronthalf

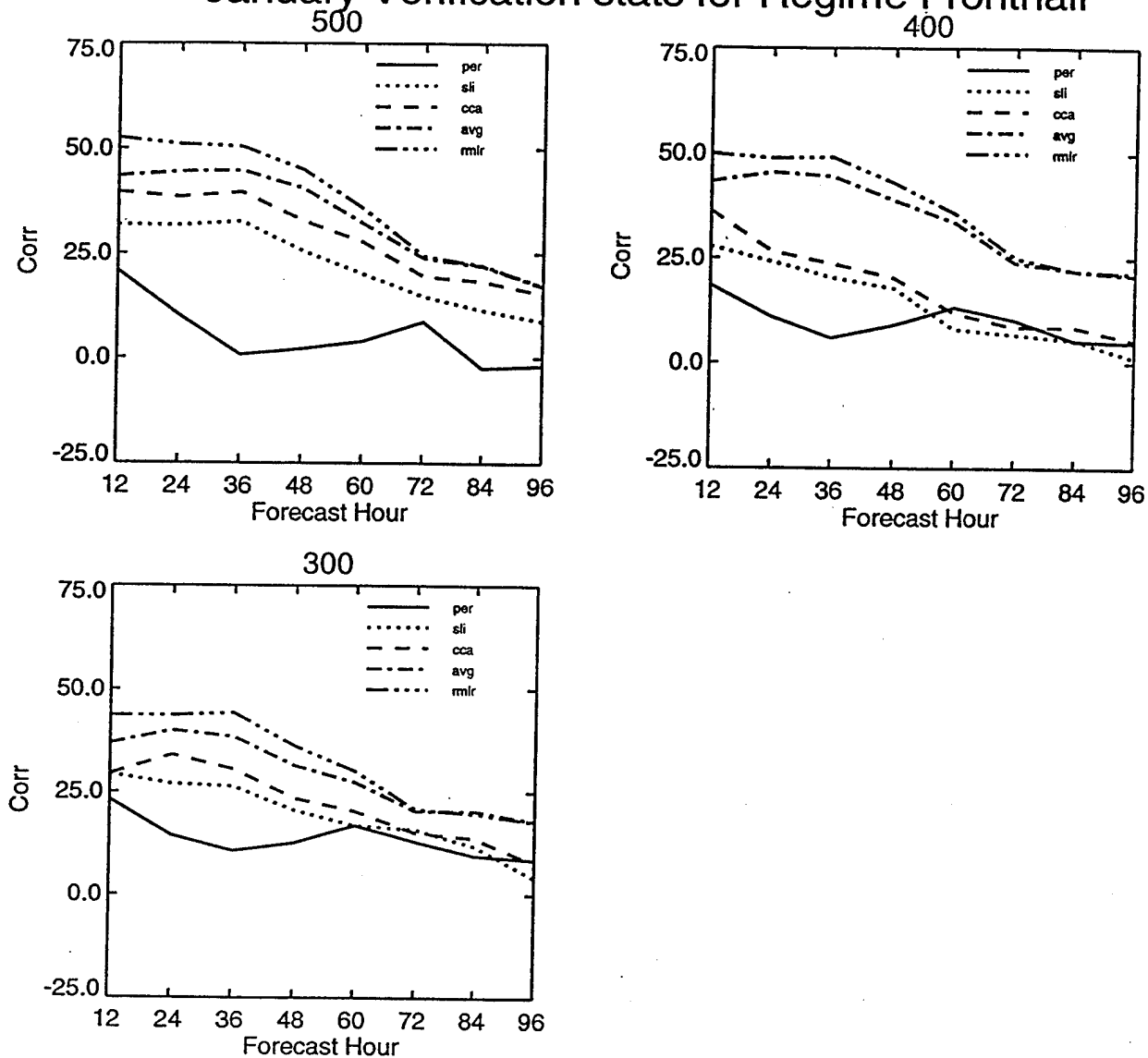


Figure 65: Continued.

## January Verification stats for Regime Backhalf

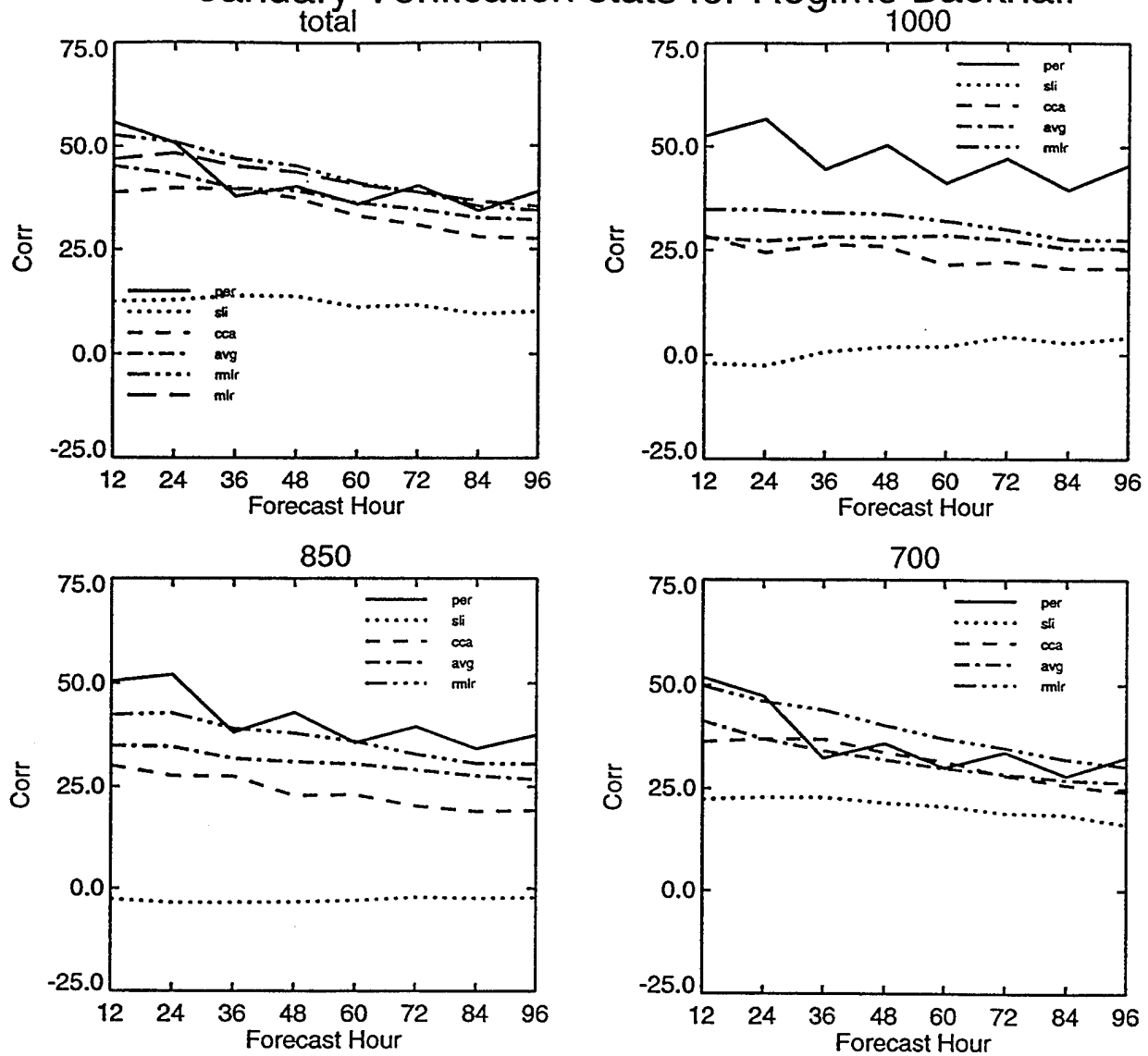


Figure 66: Same as Figure 58, except for the correlation of 12-96 hour forecasts of cloud cover, computed over the "backhalf" portion of the NH RTNEPH octagon.

# January Verification stats for Regime Backhalf

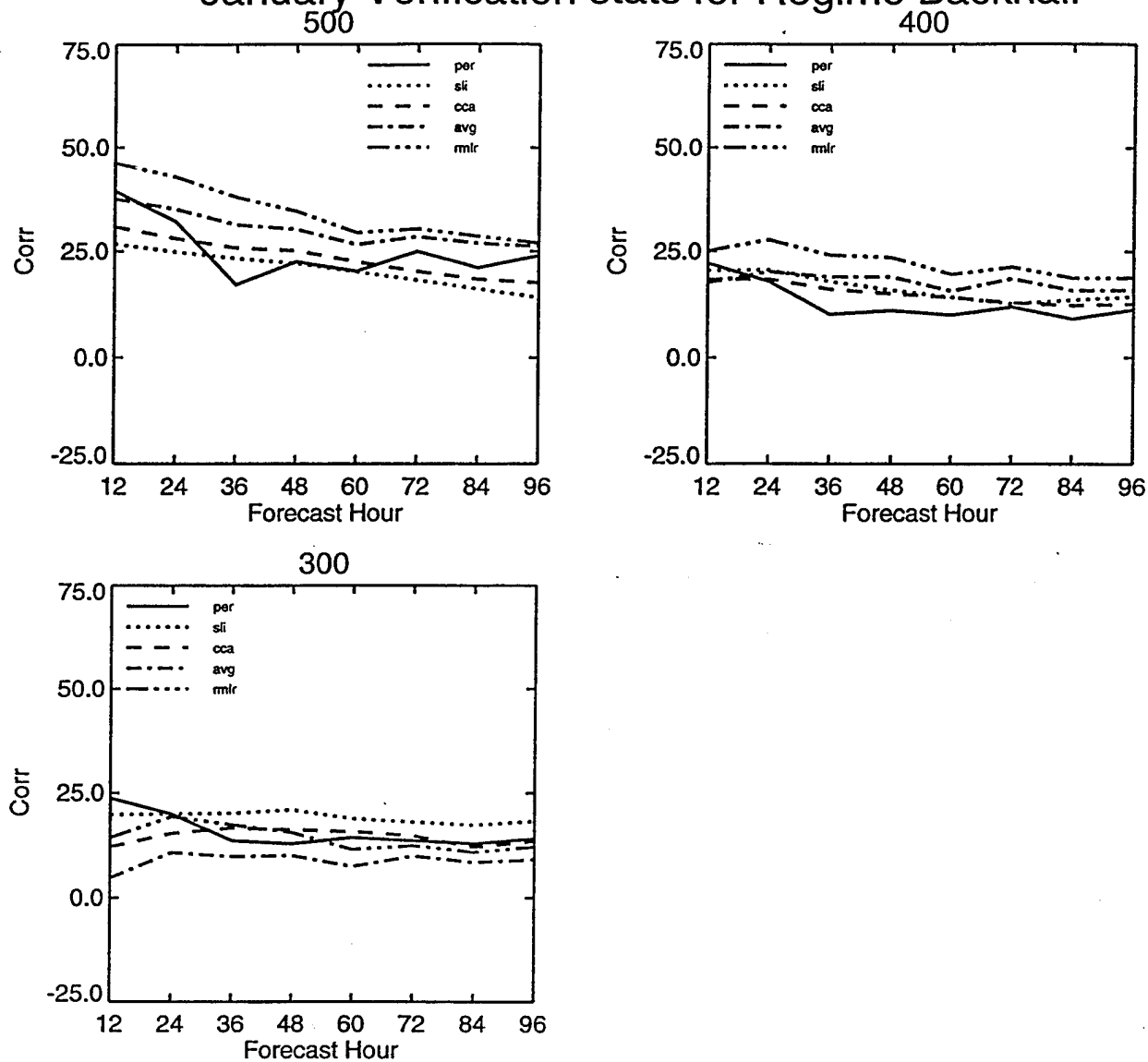


Figure 66: Continued.

# January Verification stats for Regime RTNEPH

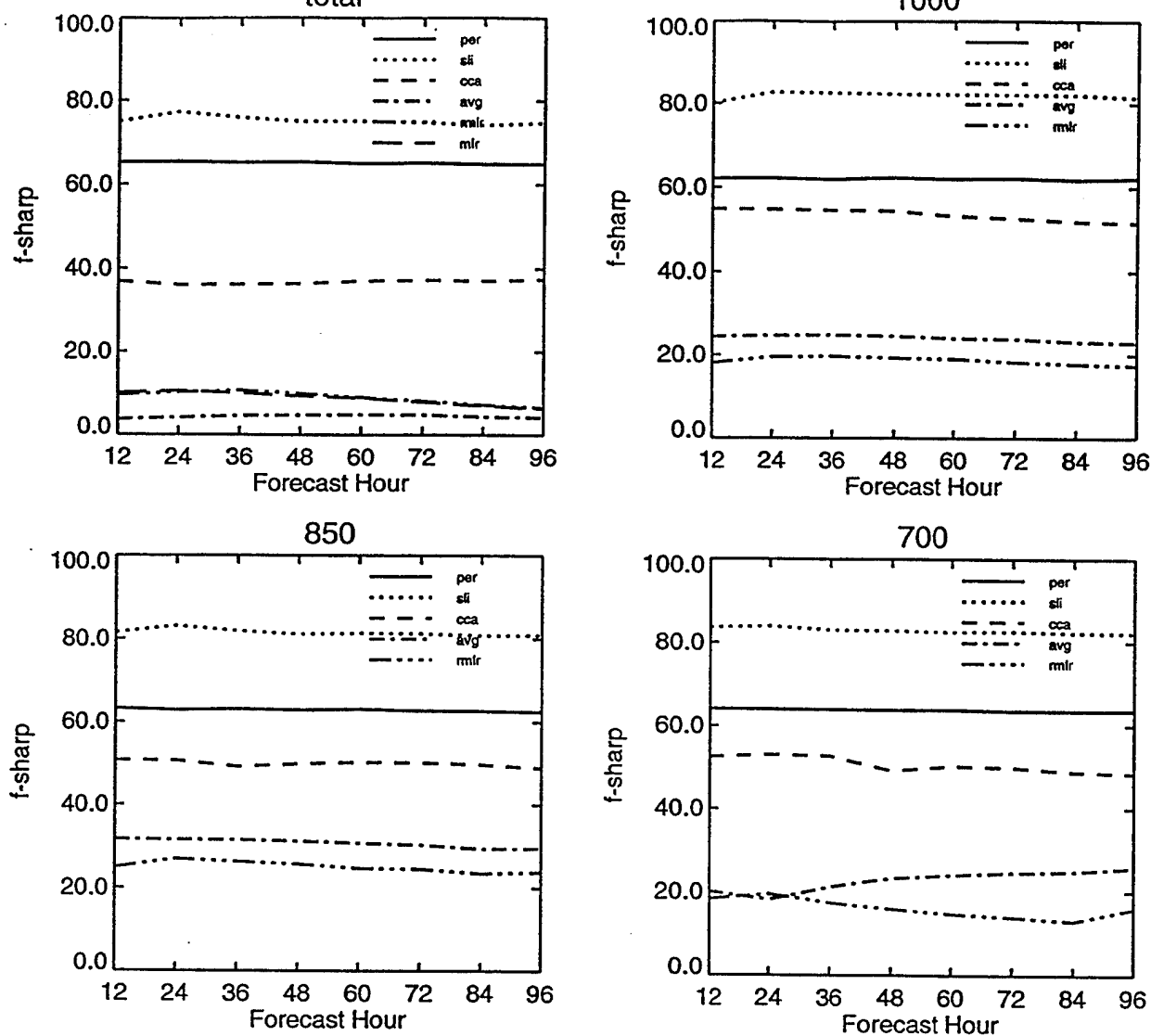


Figure 67: Same as Figure 58, except for the sharpness of 12-96 hour forecasts of cloud cover, computed over the NH RTNEPH octagon.

# January Verification stats for Regime RTNEPH

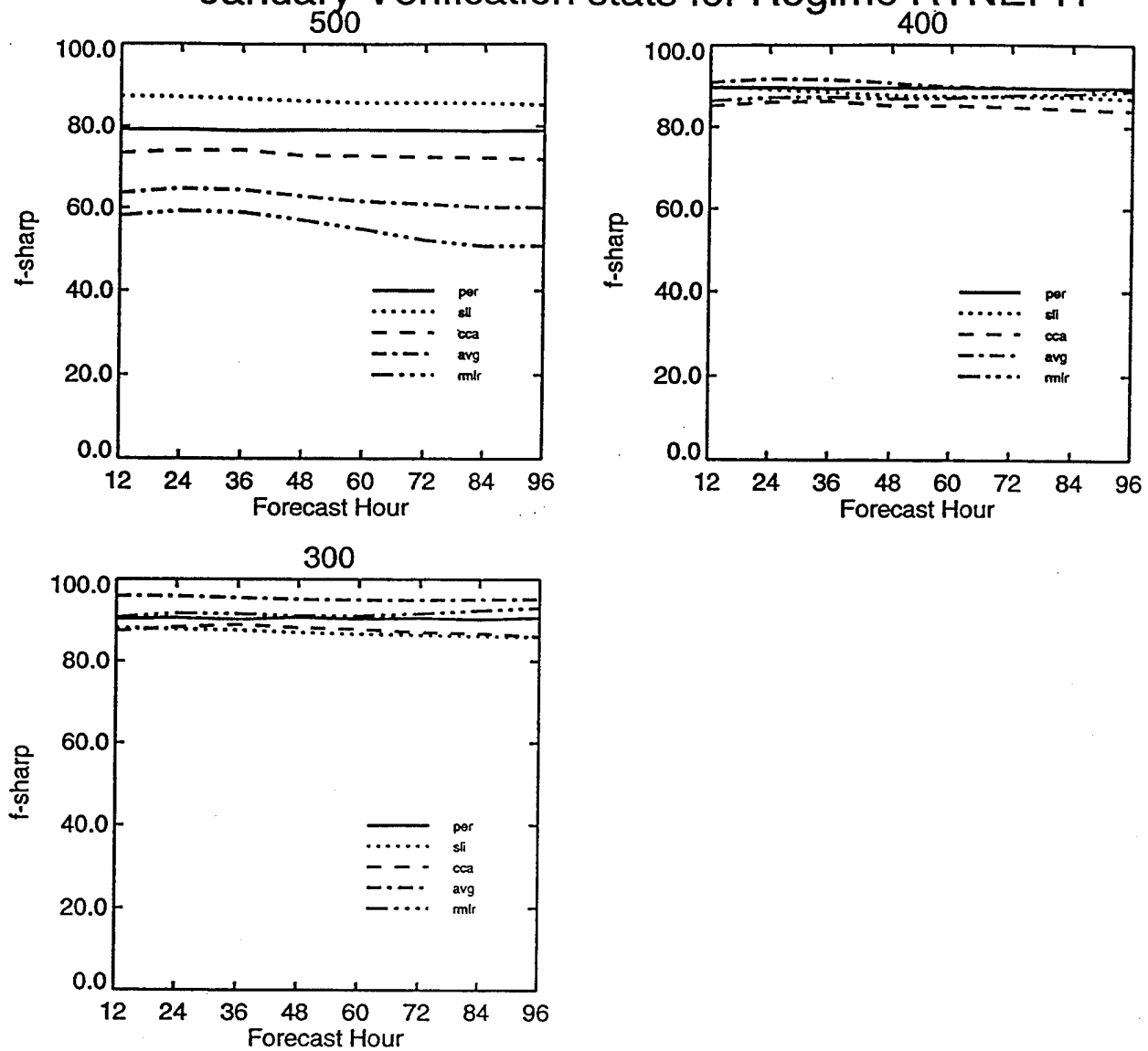


Figure 67: Continued.

# January Verification stats for Regime Fronthalf

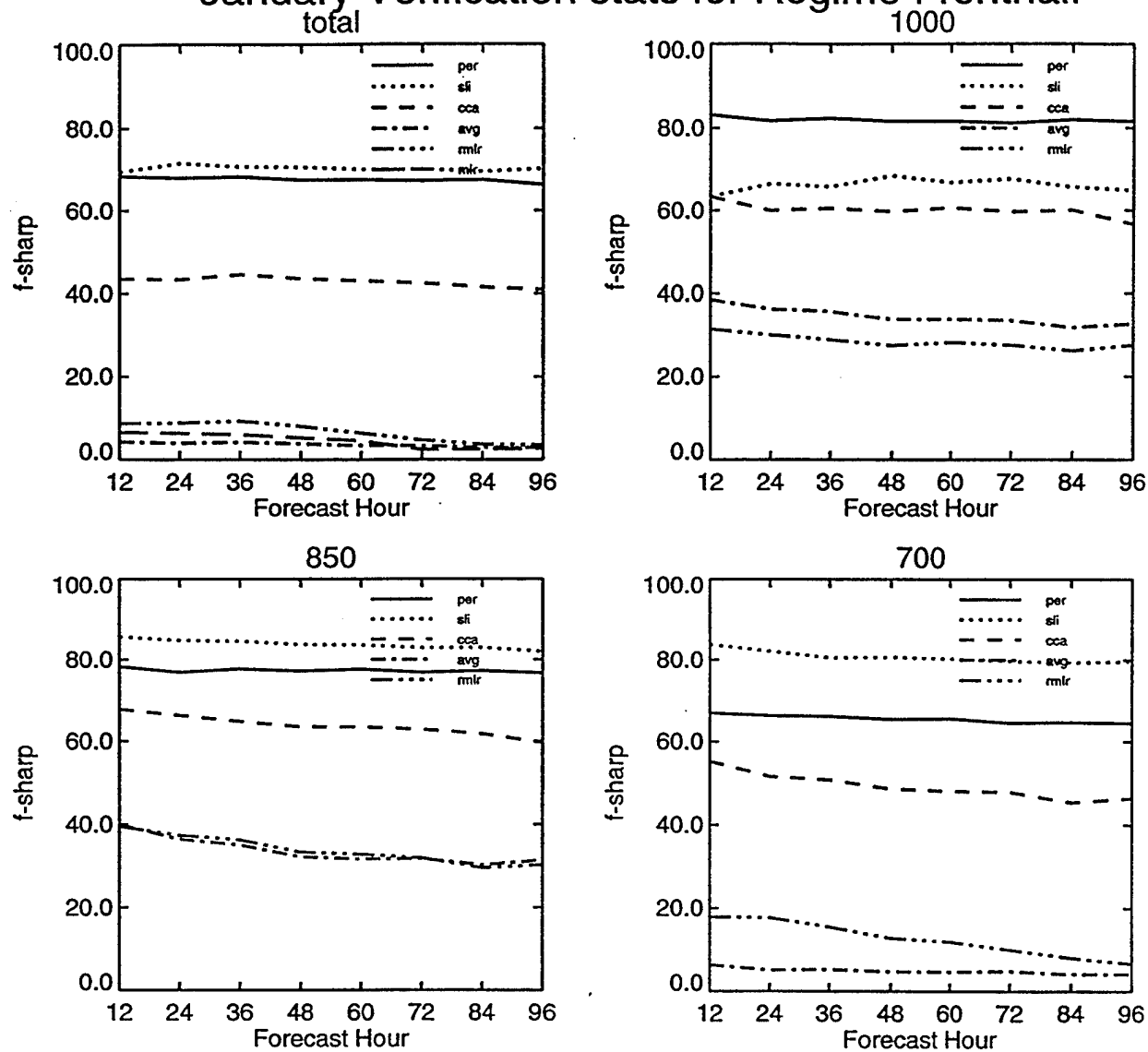


Figure 68: Same as Figure 58, except for the sharpness of 12-96 hour forecasts of cloud cover, computed over the "fronthalf" portion of the NH RTNEPH octagon.

# January Verification stats for Regime Fronthalf

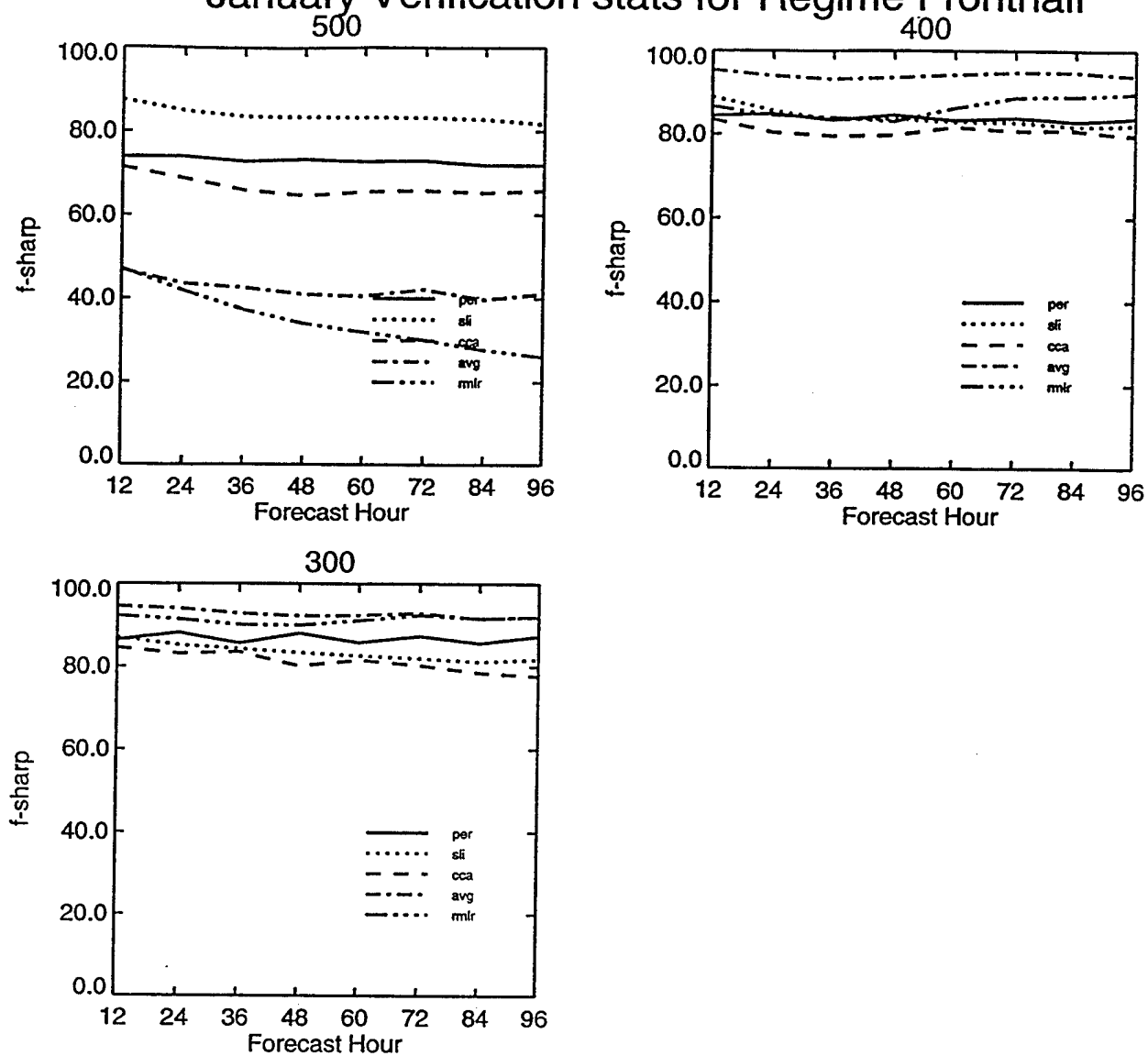


Figure 68: Continued.

## January Verification stats for Regime Backhalf

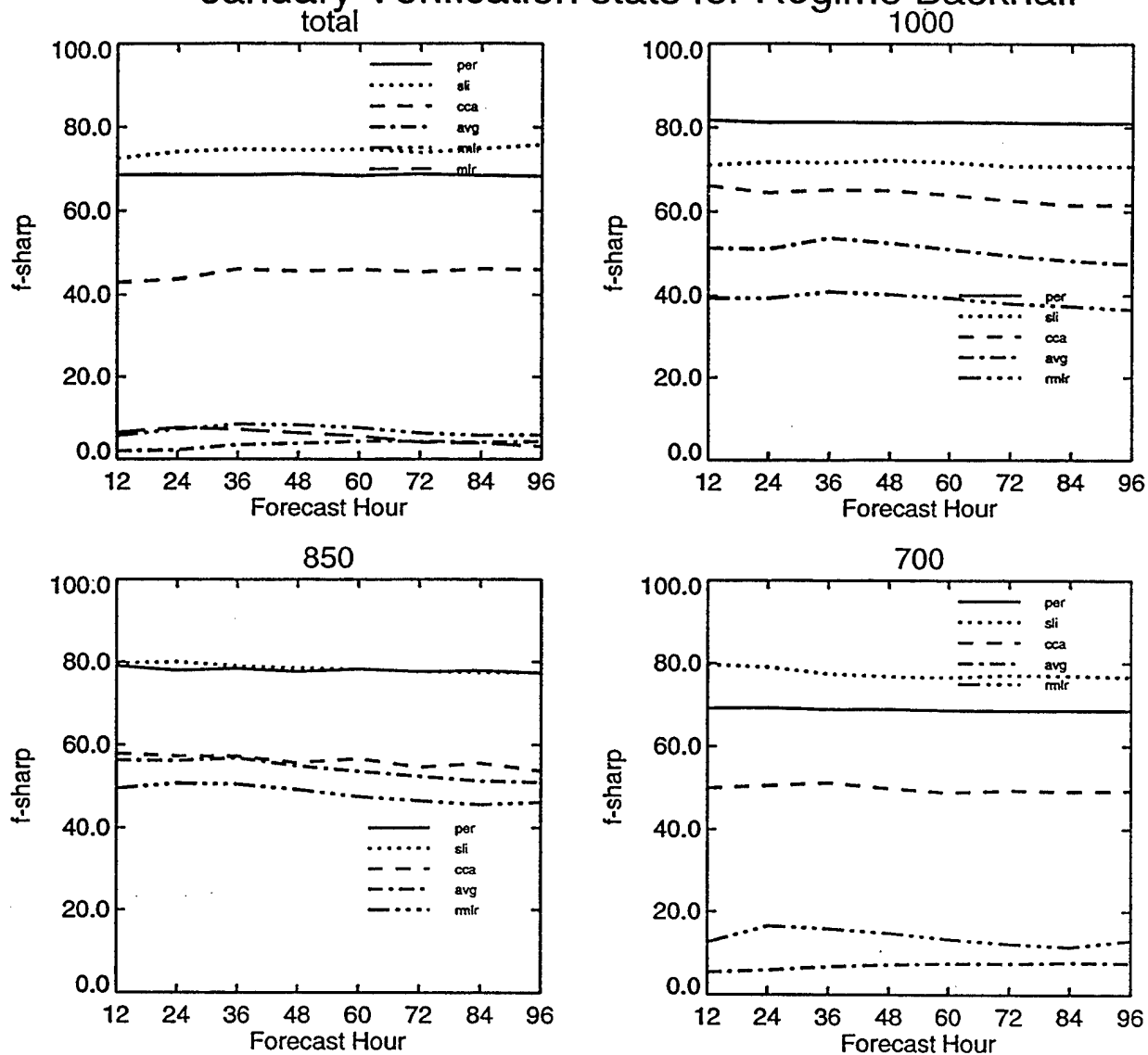


Figure 69: Same as Figure 58, except for the sharpness of 12-96 hour forecasts of cloud cover, computed over the "backhalf" portion of the NH RTNEPH octagon.



# January Verification stats for Regime Backhalf

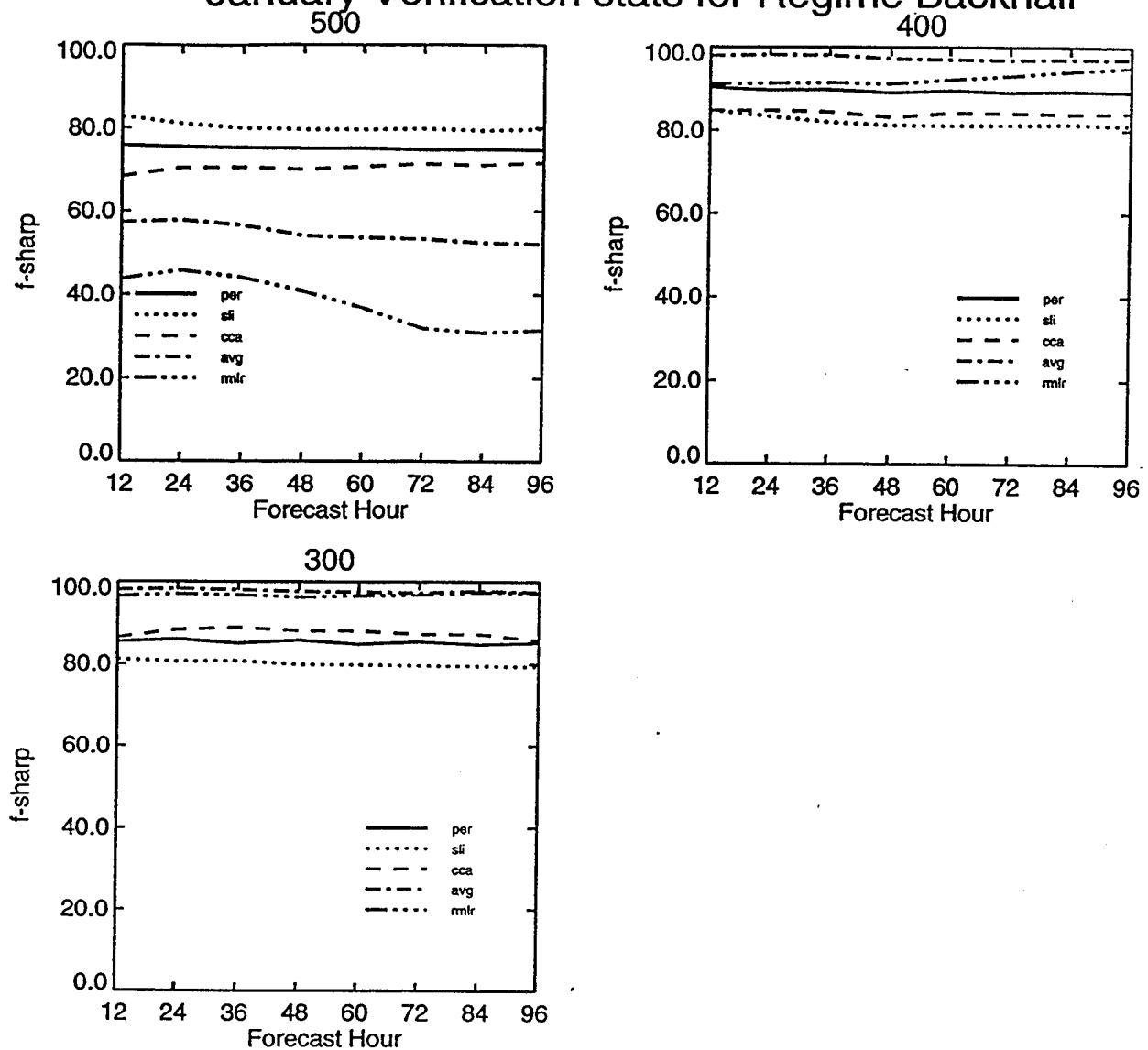


Figure 69: Continued.

## April Verification stats for Regime RTNEPH

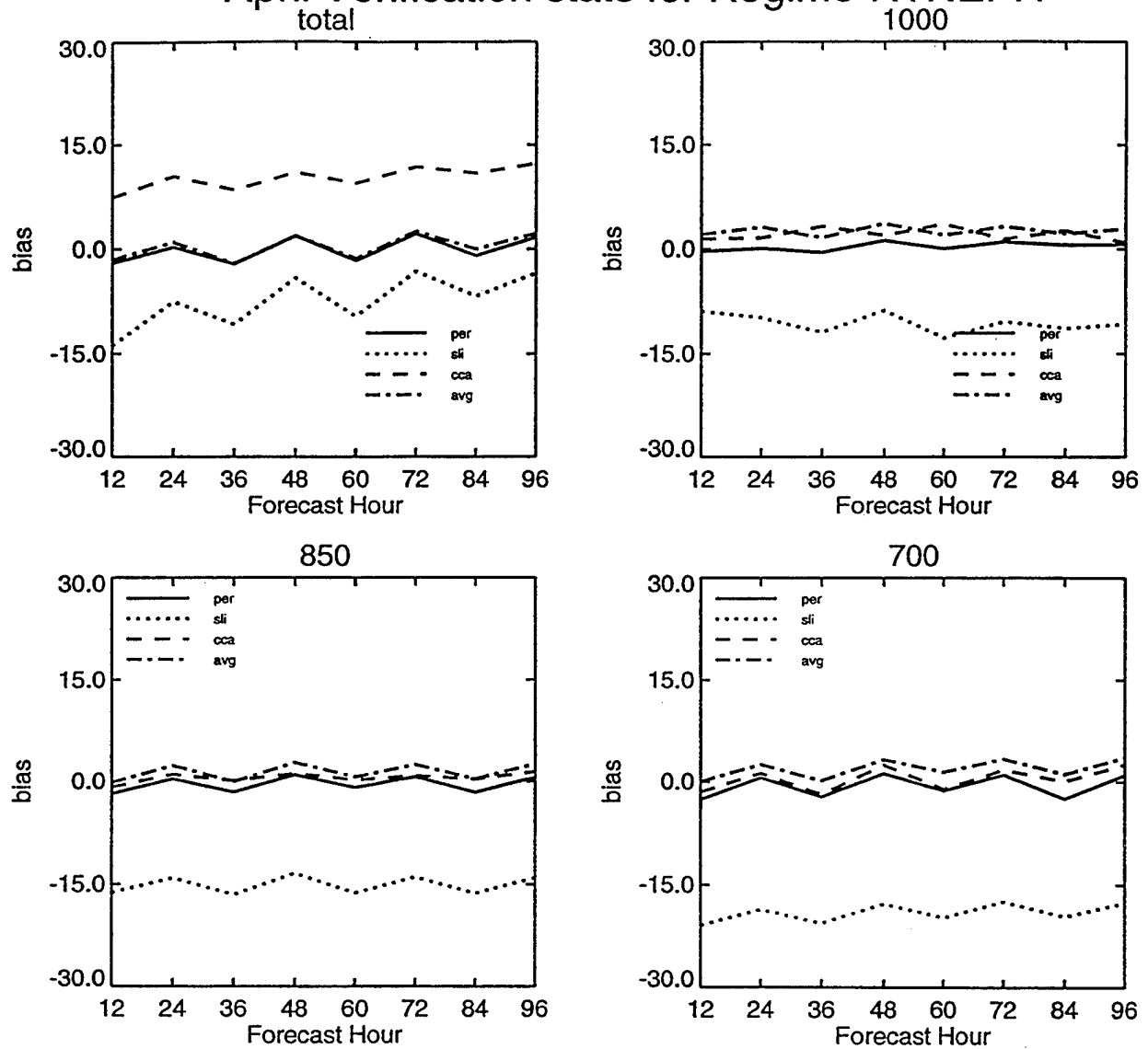


Figure 70: Same as Figure 58, except for the bias of 12-96 hour forecasts of cloud cover for April, computed over the NH RTNEPH octagon.

# April Verification stats for Regime RTNEPH

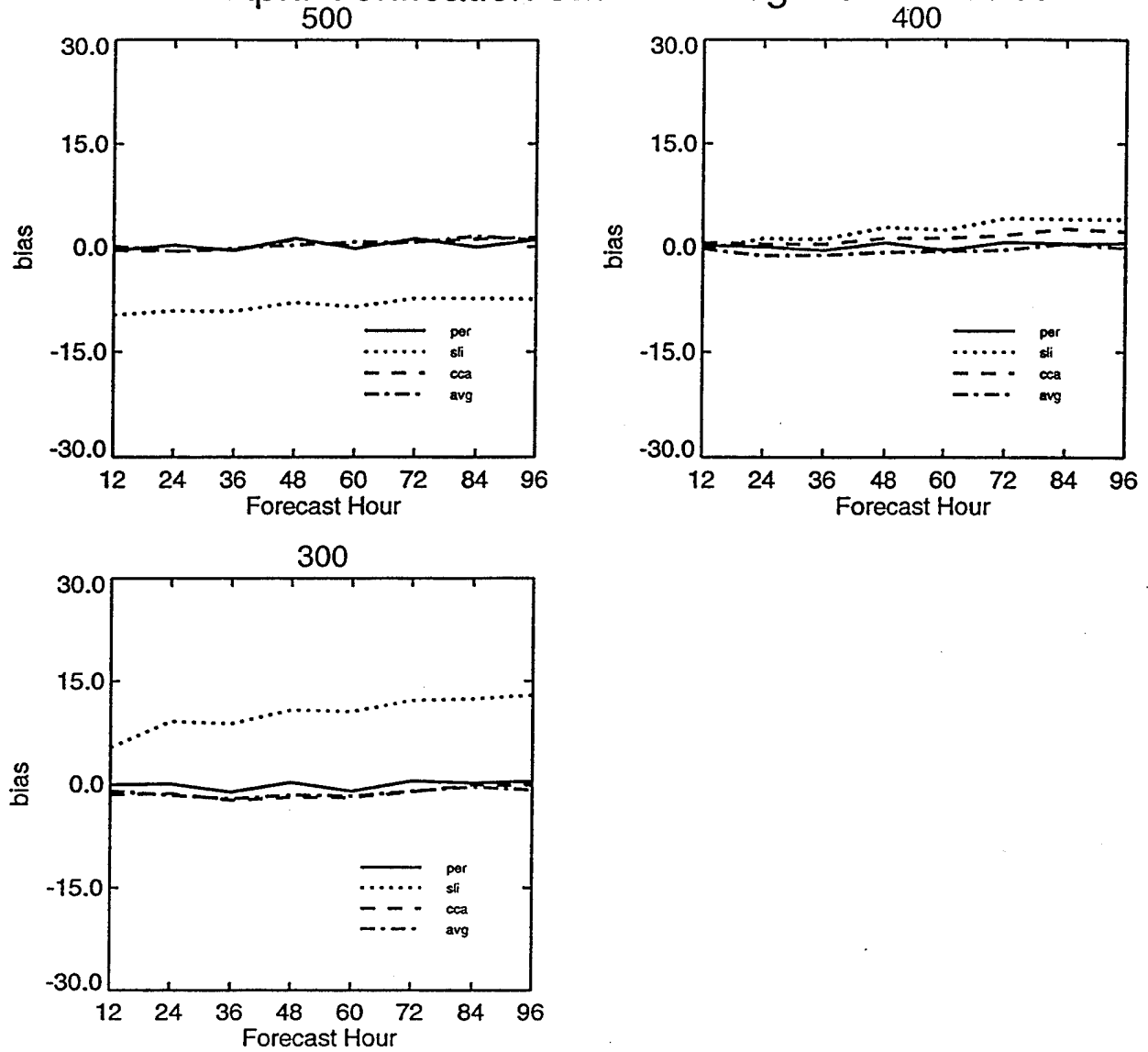


Figure 70: Continued.

# July Verification stats for Regime RTNEPH

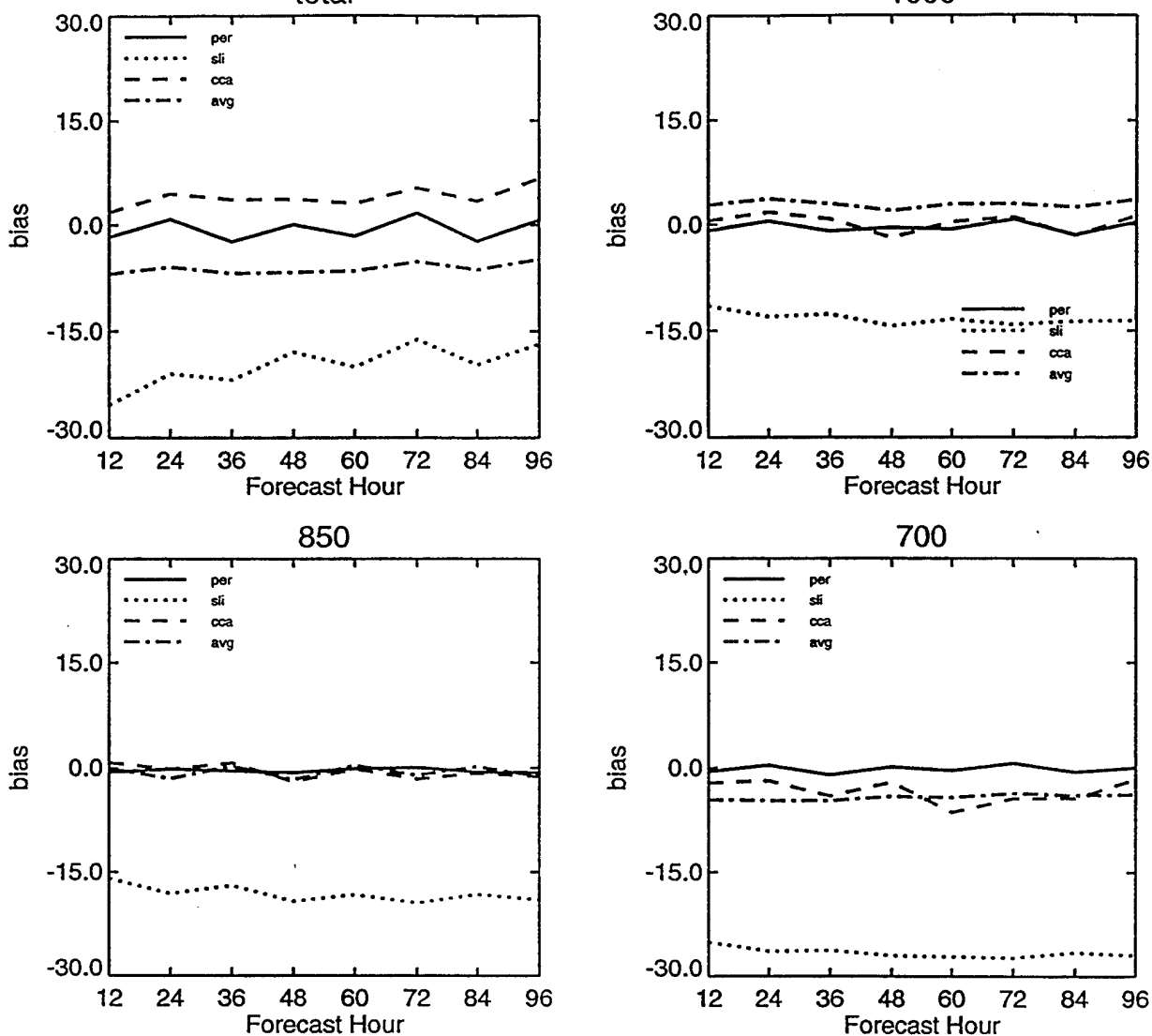


Figure 71: Same as Figure 58, except for the bias of 12-96 hour forecasts of cloud cover for July, computed over the NH RTNEPH octagon.

# July Verification stats for Regime RTNEPH

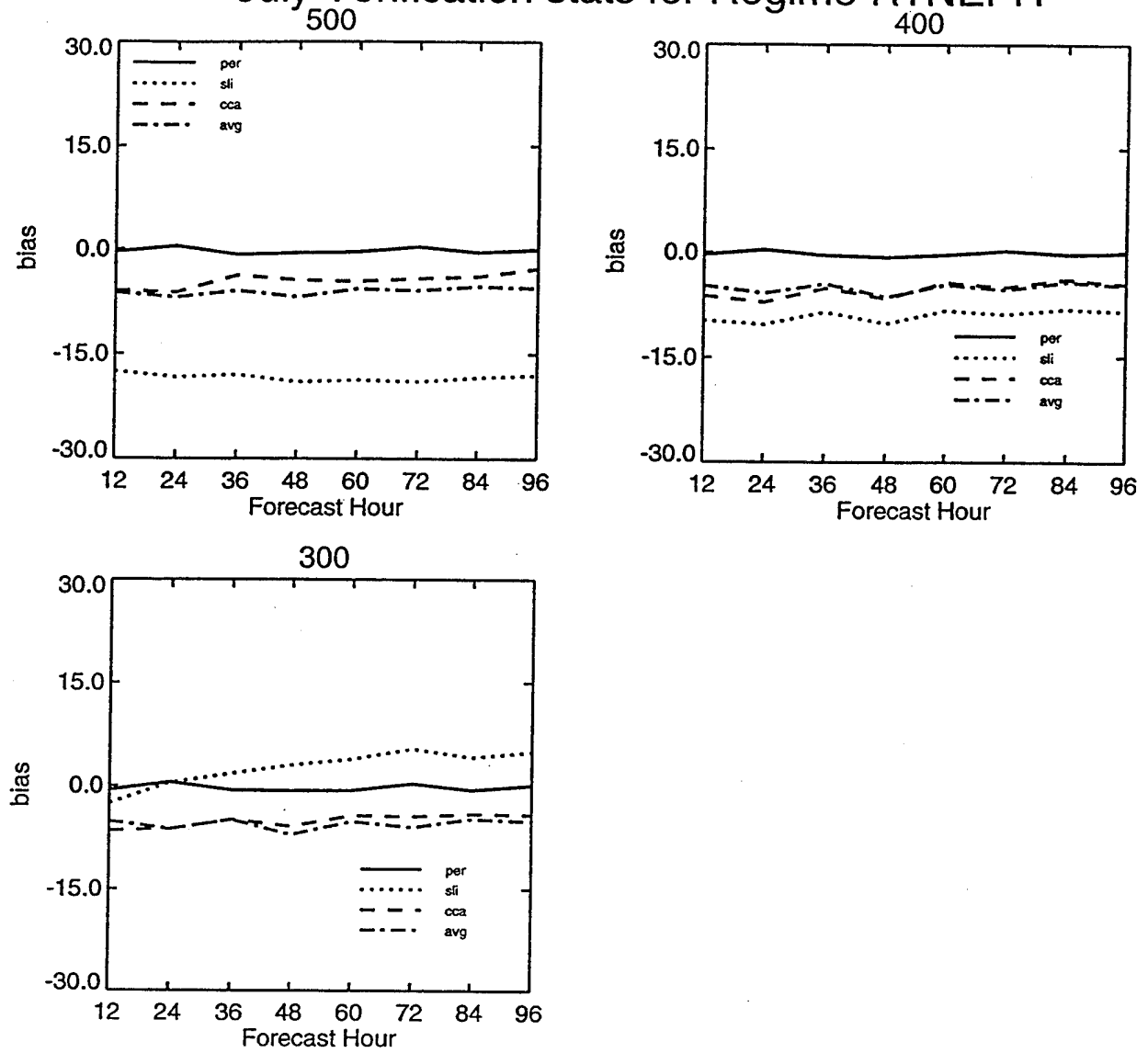


Figure 71: Continued.

## October Verification stats for Regime RTNEPH

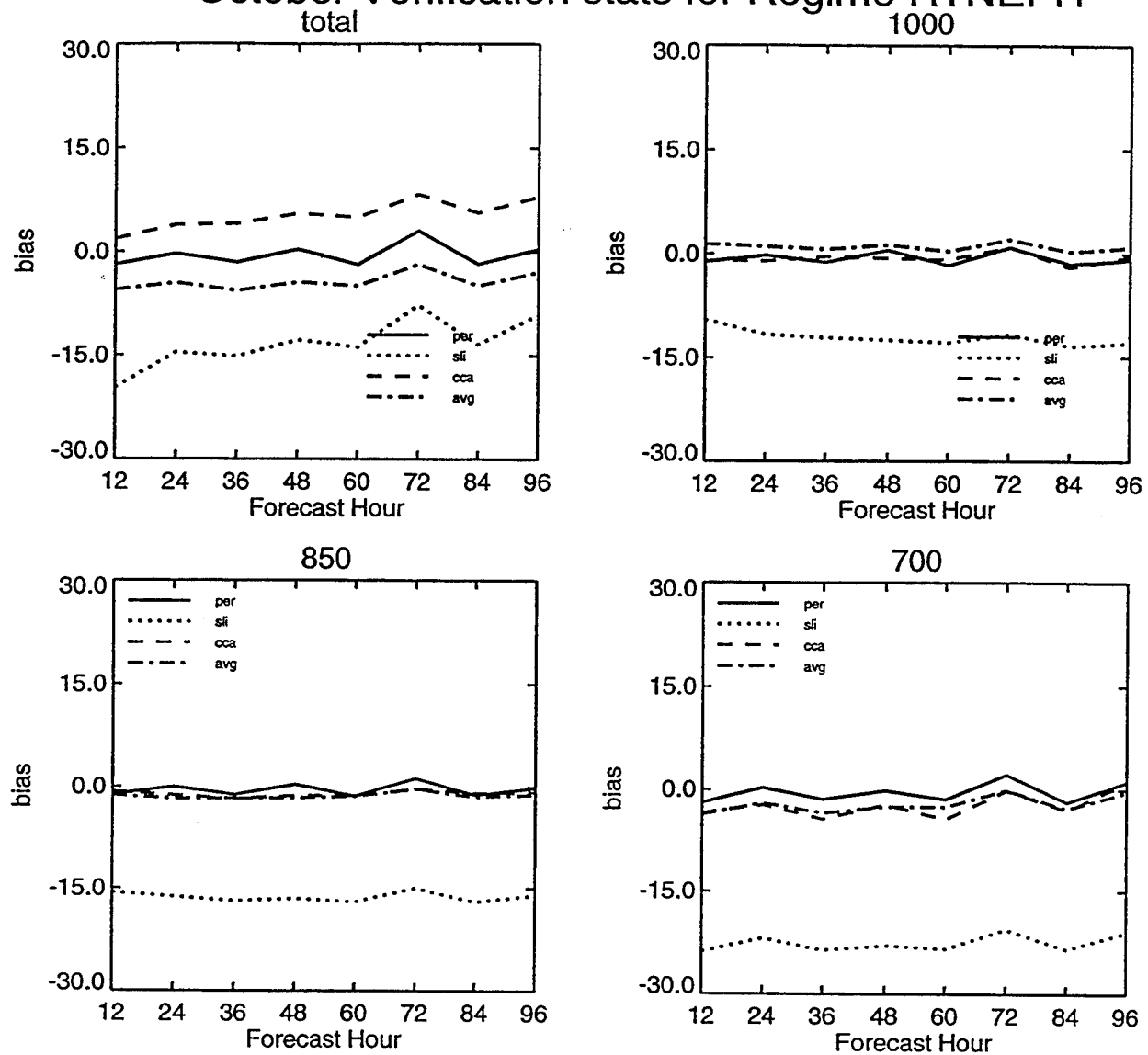


Figure 72: Same as Figure 58, except for the bias of 12-96 hour forecasts of cloud cover for October, computed over the NH RTNEPH octagon.

# October Verification stats for Regime RTNEPH

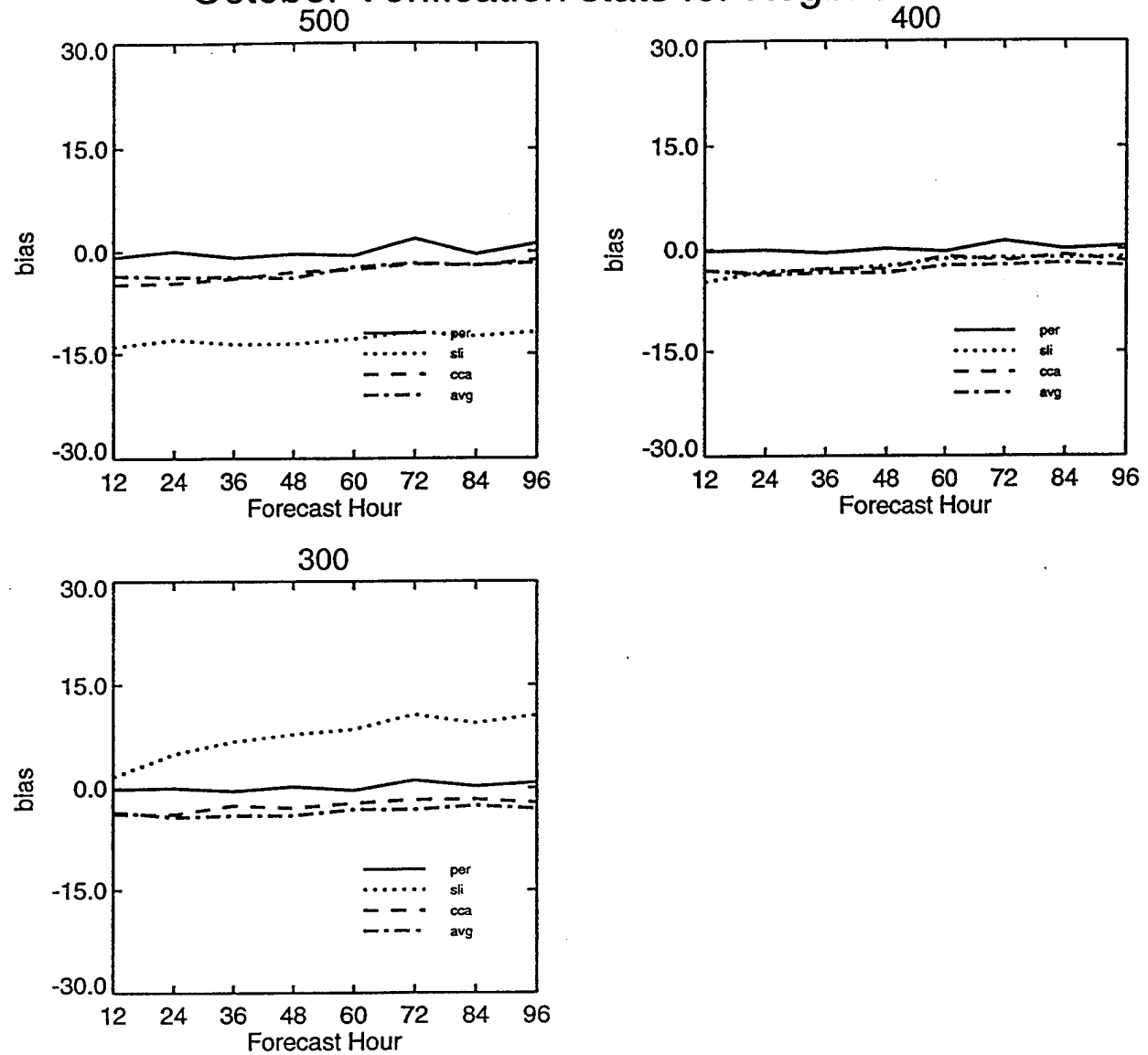


Figure 72: Continued.

# April Verification stats for Regime RTNEPH

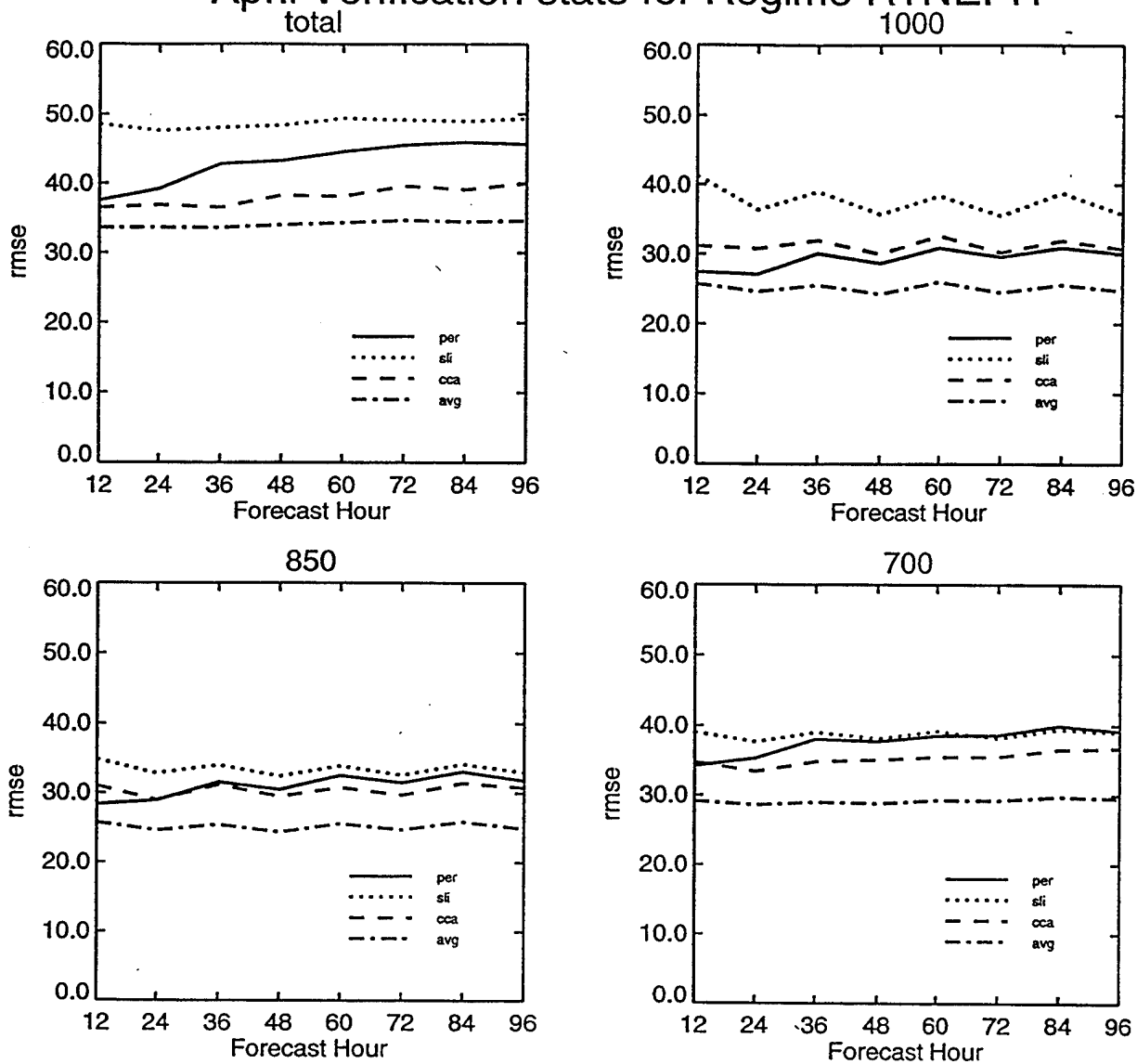


Figure 73: Same as Figure 58, except for the rmse of 12-96 hour forecasts of cloud cover for April, computed over the NH RTNEPH octagon.



# April Verification stats for Regime RTNEPH

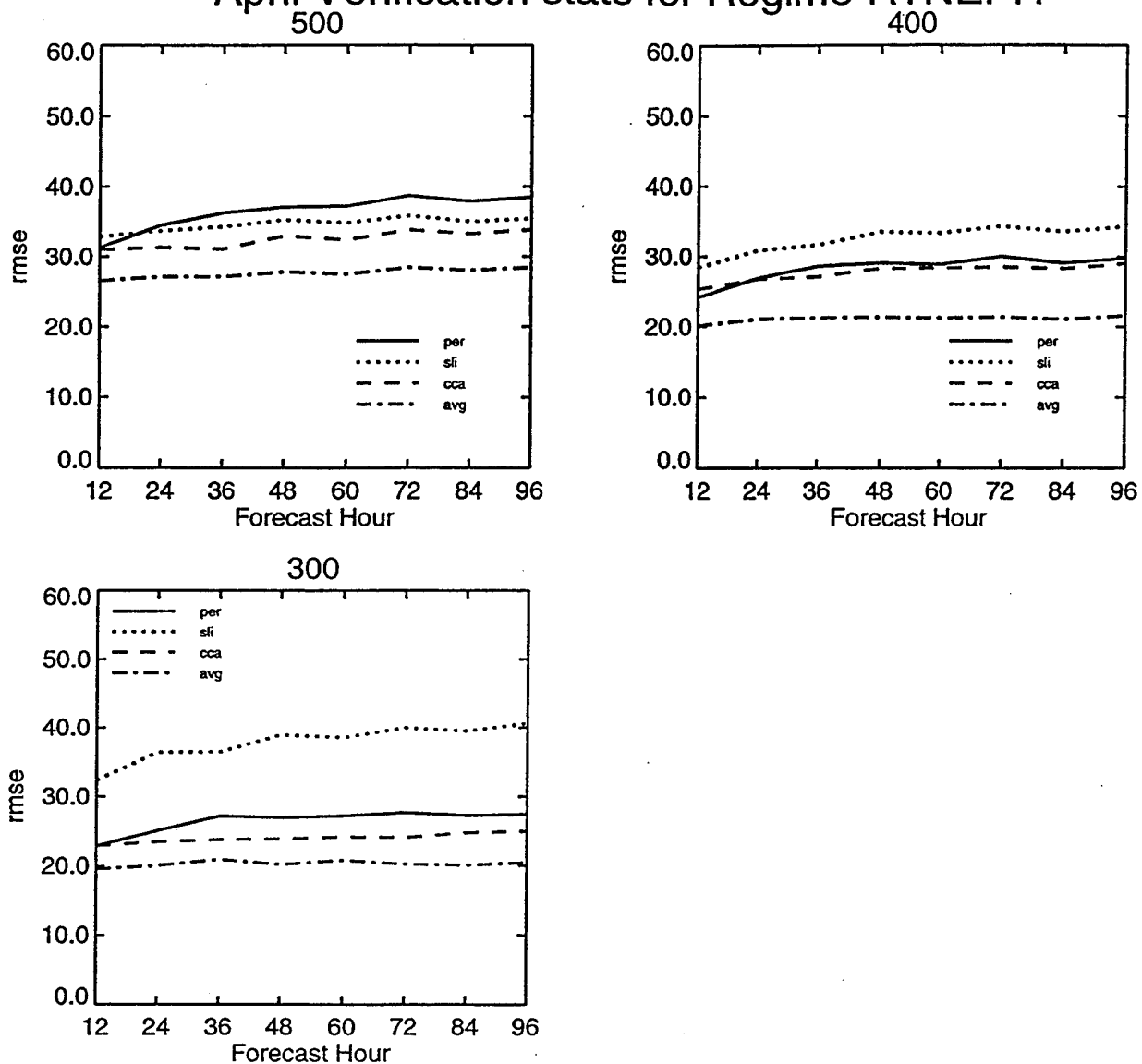


Figure 73: Continued.

## July Verification stats for Regime RTNEPH

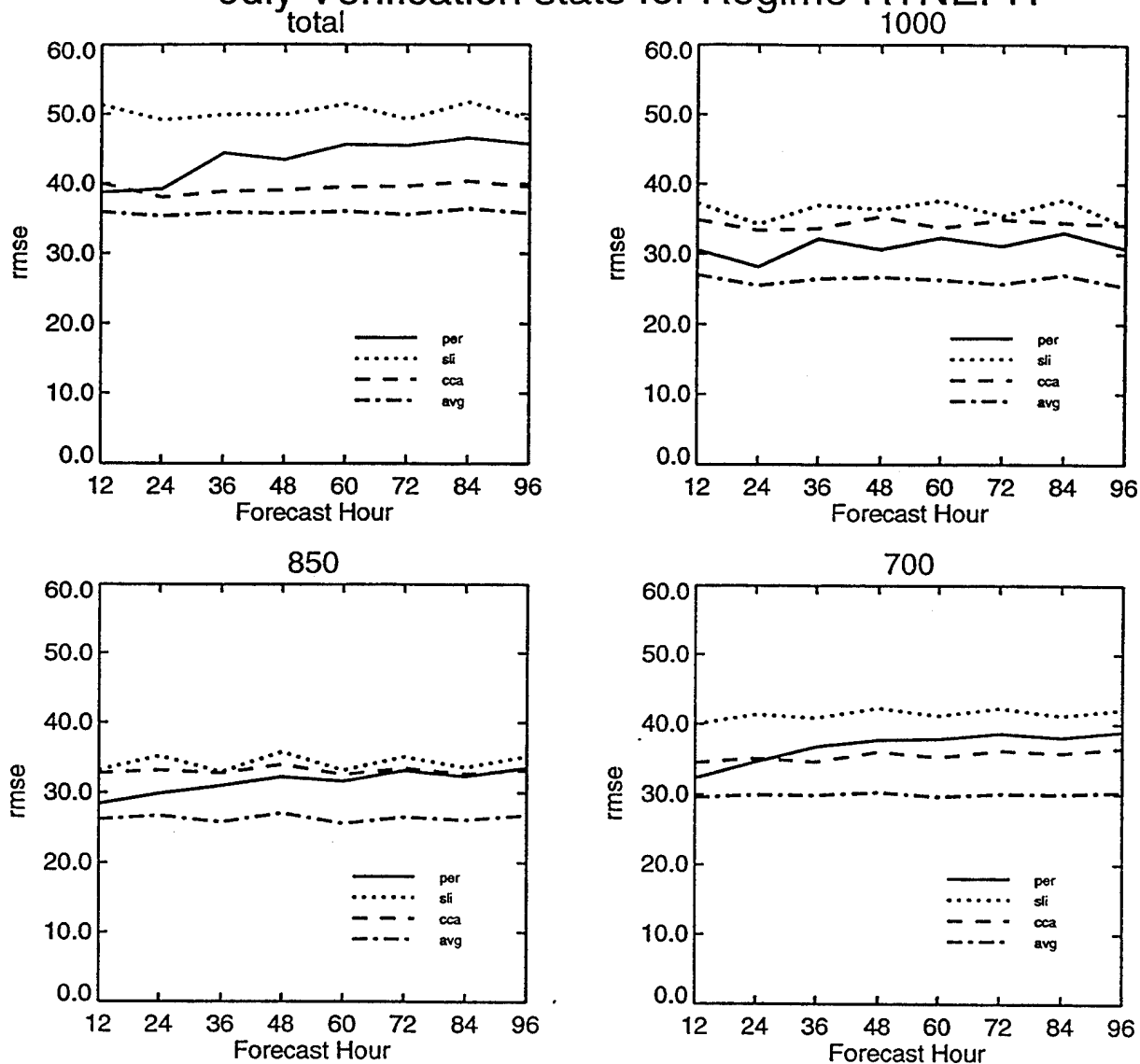


Figure 74: Same as Figure 58, except for the rmse of 12-96 hour forecasts of cloud cover for July, computed over the NH RTNEPH octagon.

# July Verification stats for Regime RTNEPH

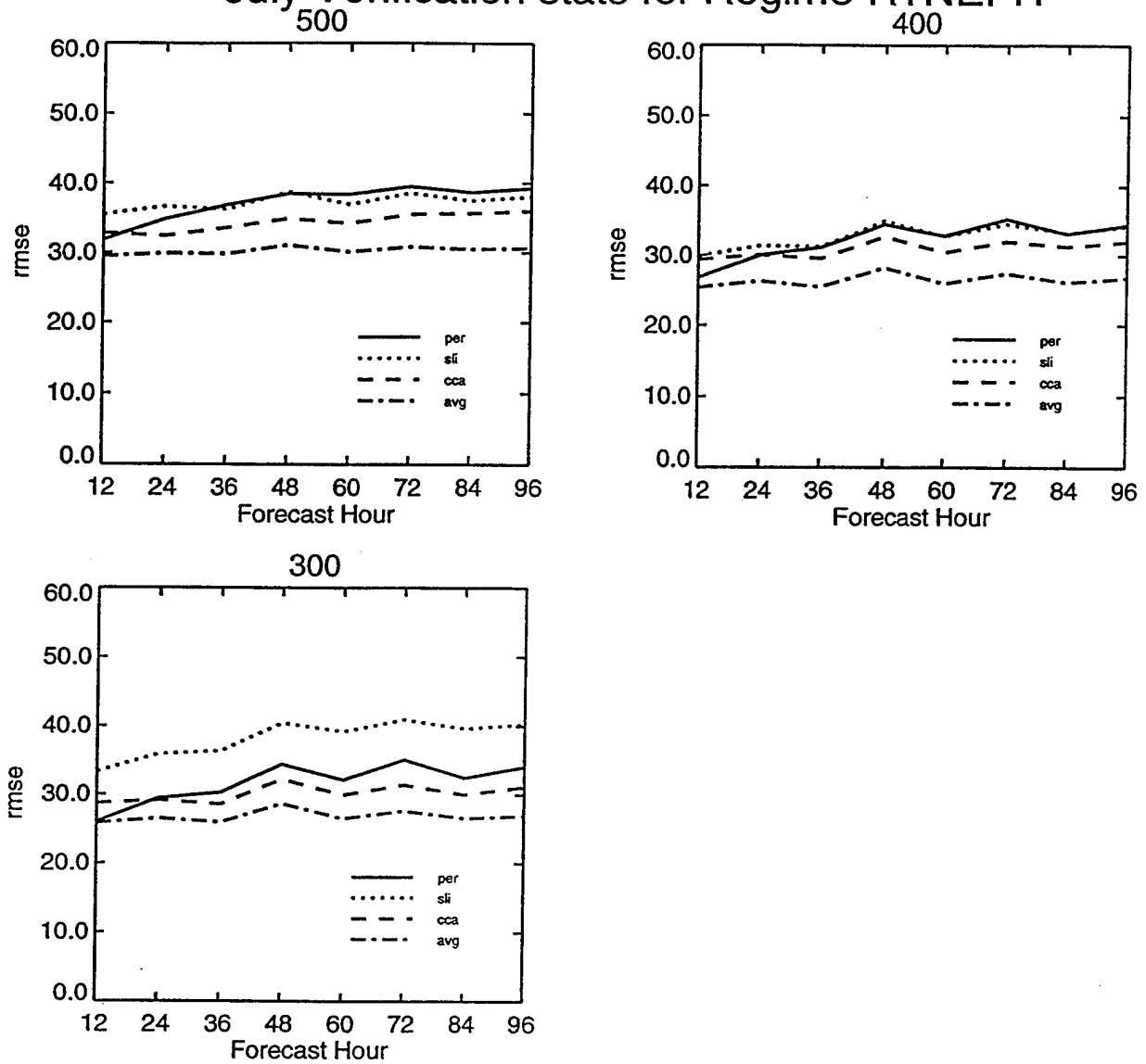


Figure 74: Continued.

## October Verification stats for Regime RTNEPH

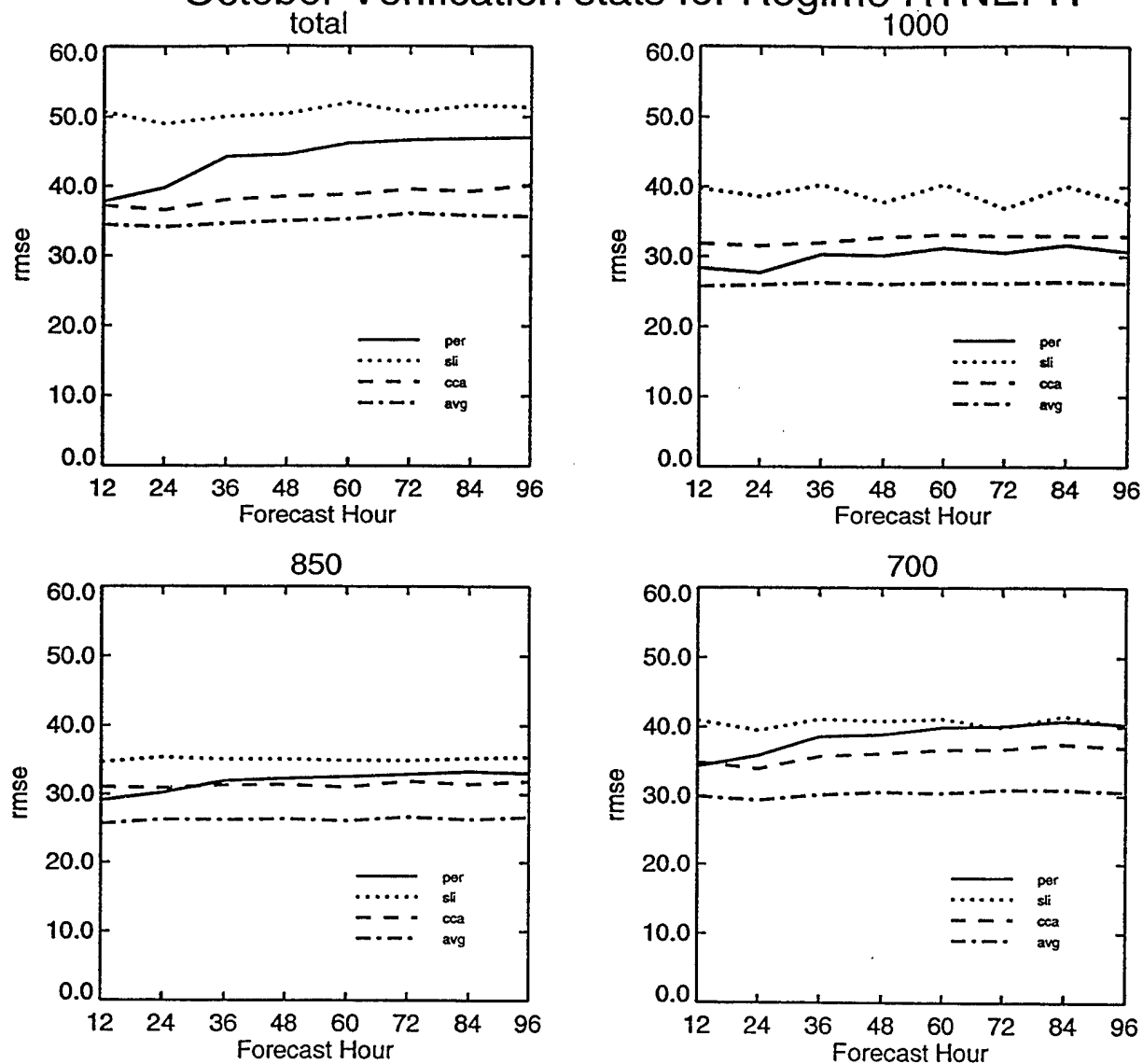


Figure 75: Same as Figure 58, except for the rmse of 12-96 hour forecasts of cloud cover for October, computed over the NH RTNEPH octagon.

# October Verification stats for Regime RTNEPH

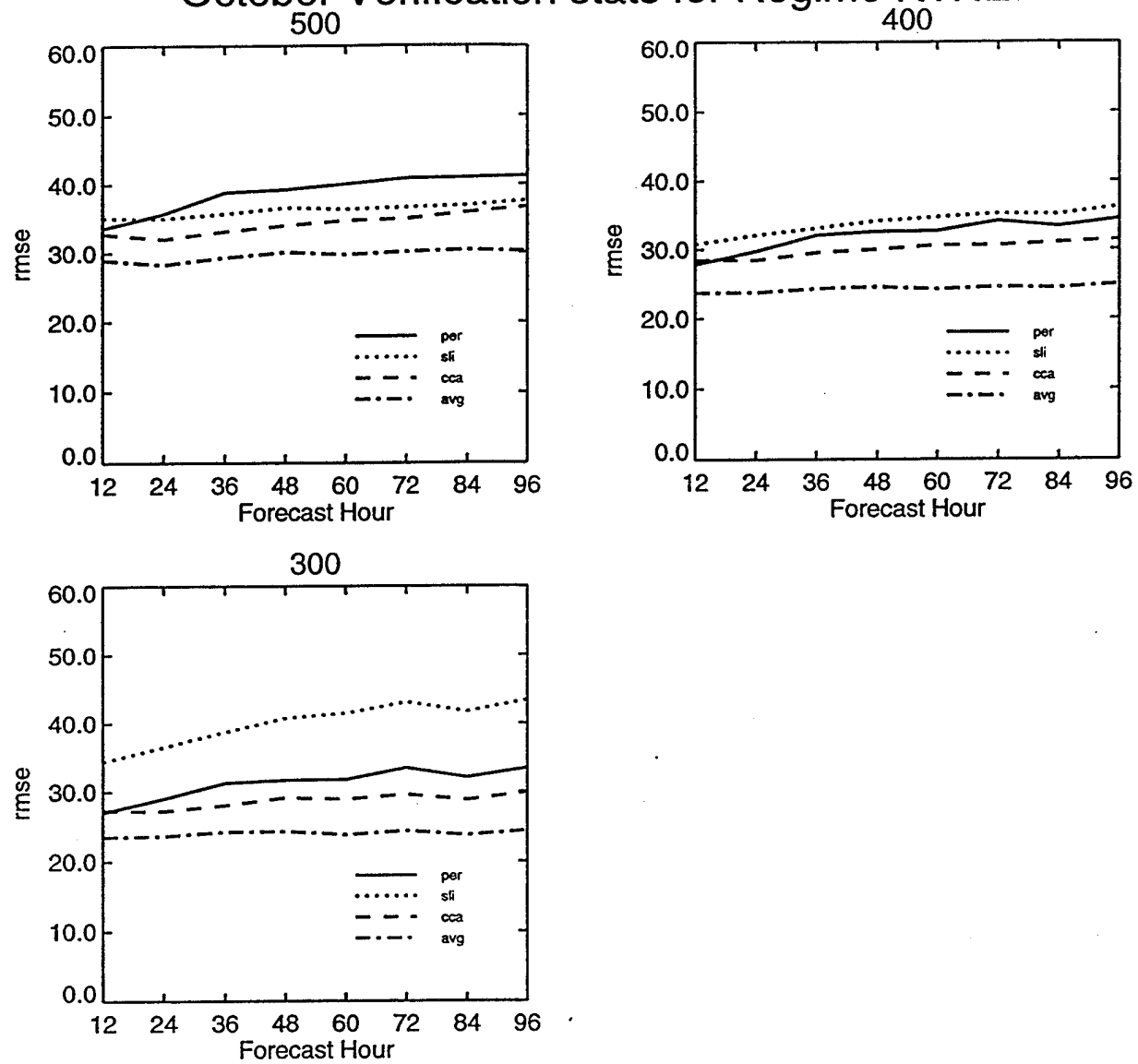


Figure 75: Continued.

# April Verification stats for Regime RTNEPH

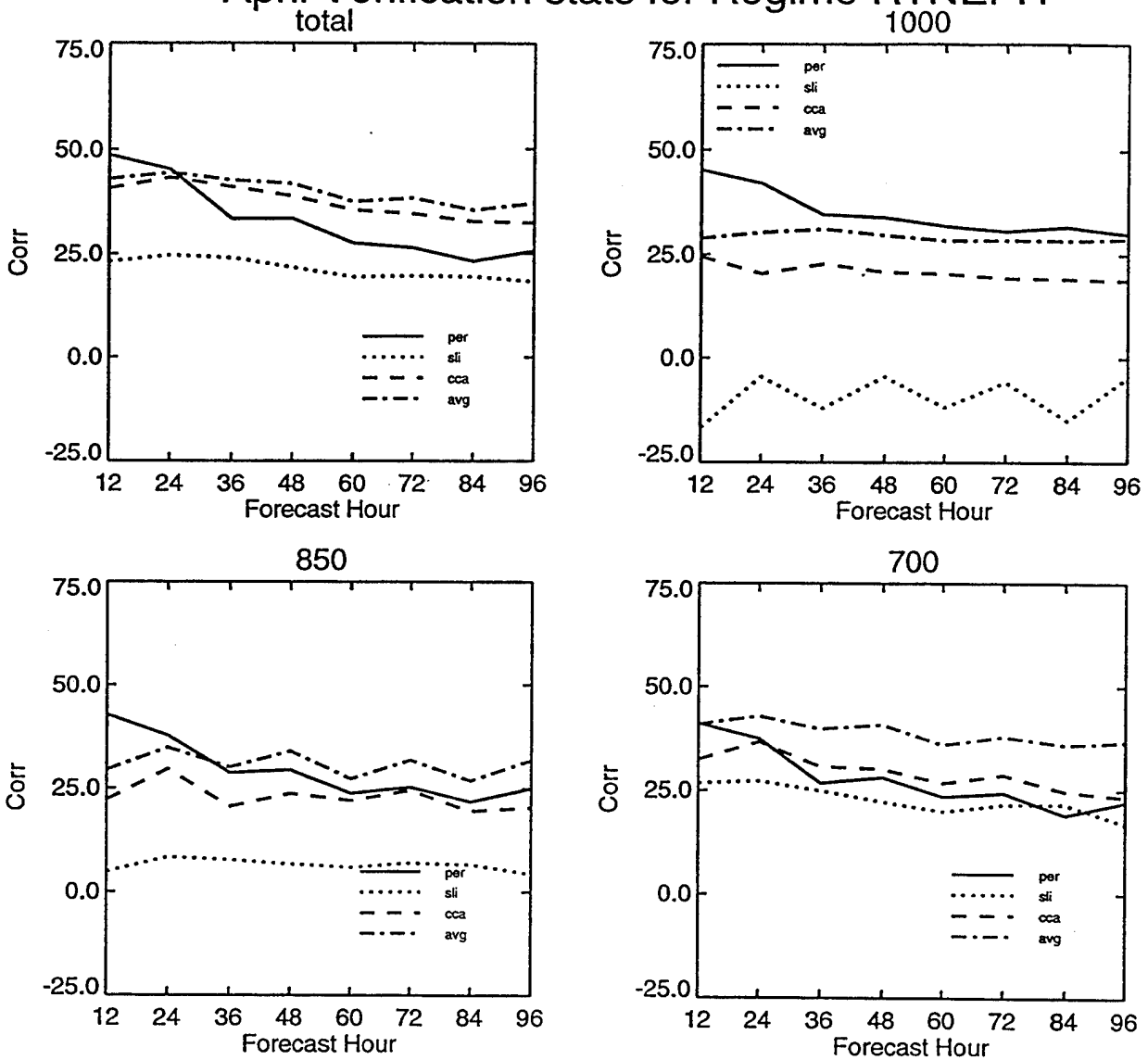


Figure 76: Same as Figure 58, except for the correlation of 12-96 hour forecasts of cloud cover for April, computed over the NH RTNEPH octagon.

# April Verification stats for Regime RTNEPH

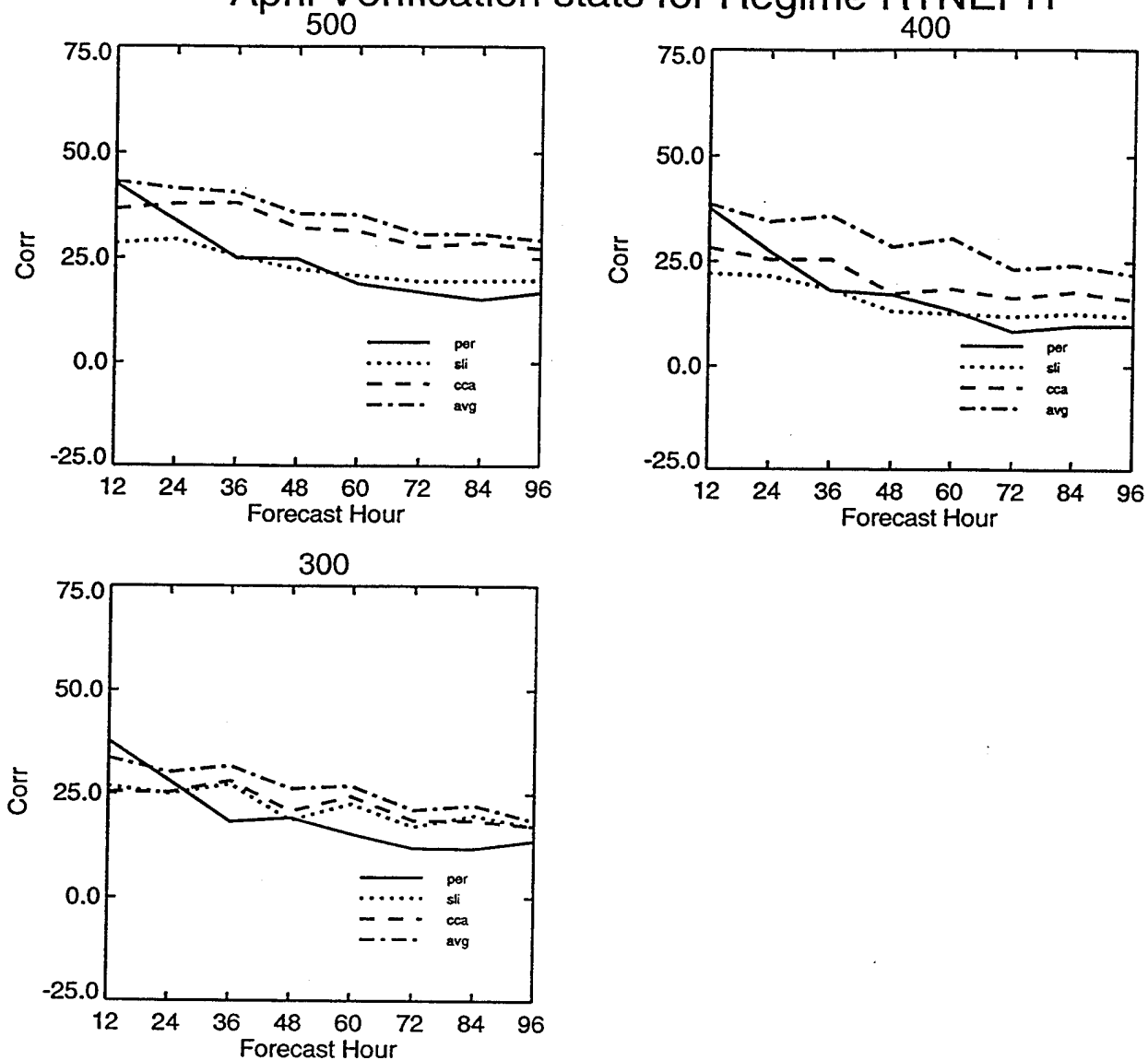


Figure 76: Continued.

## July Verification stats for Regime RTNEPH

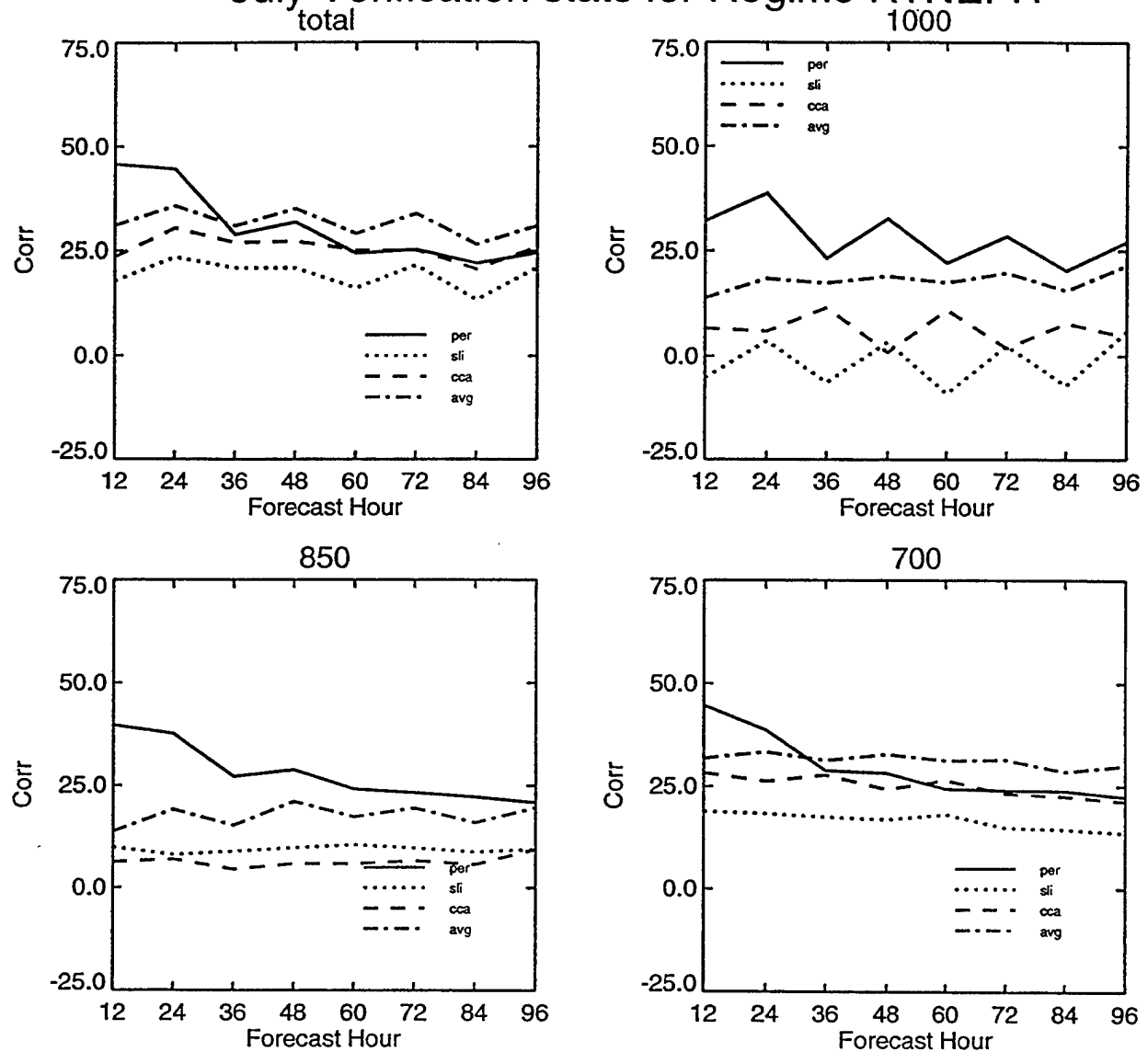


Figure 77: Same as Figure 58, except for the correlation of 12-96 hour forecasts of cloud cover for July, computed over the NH RTNEPH octagon.



# July Verification stats for Regime RTNEPH

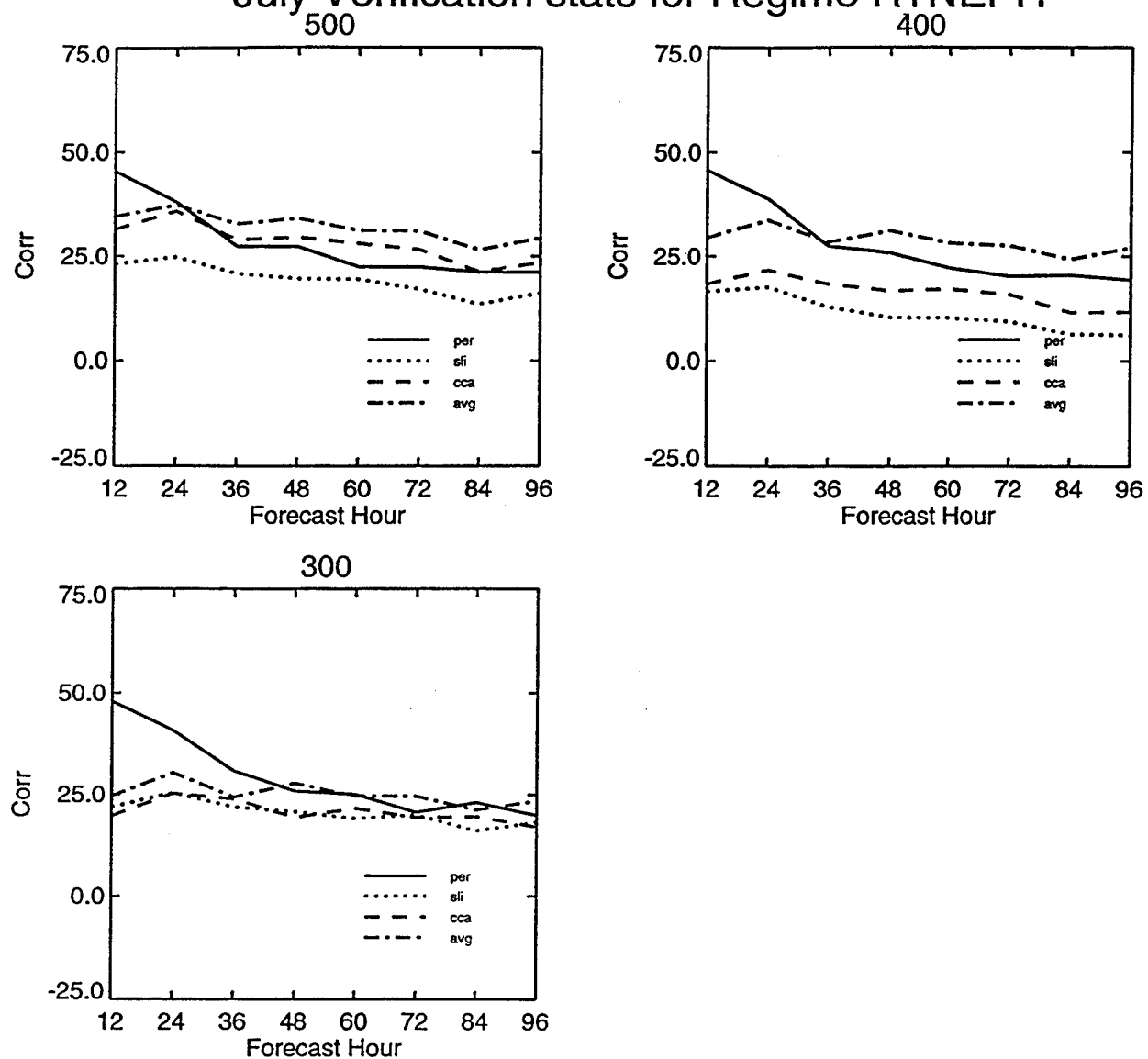


Figure 77: Continued.

## October Verification stats for Regime RTNEPH

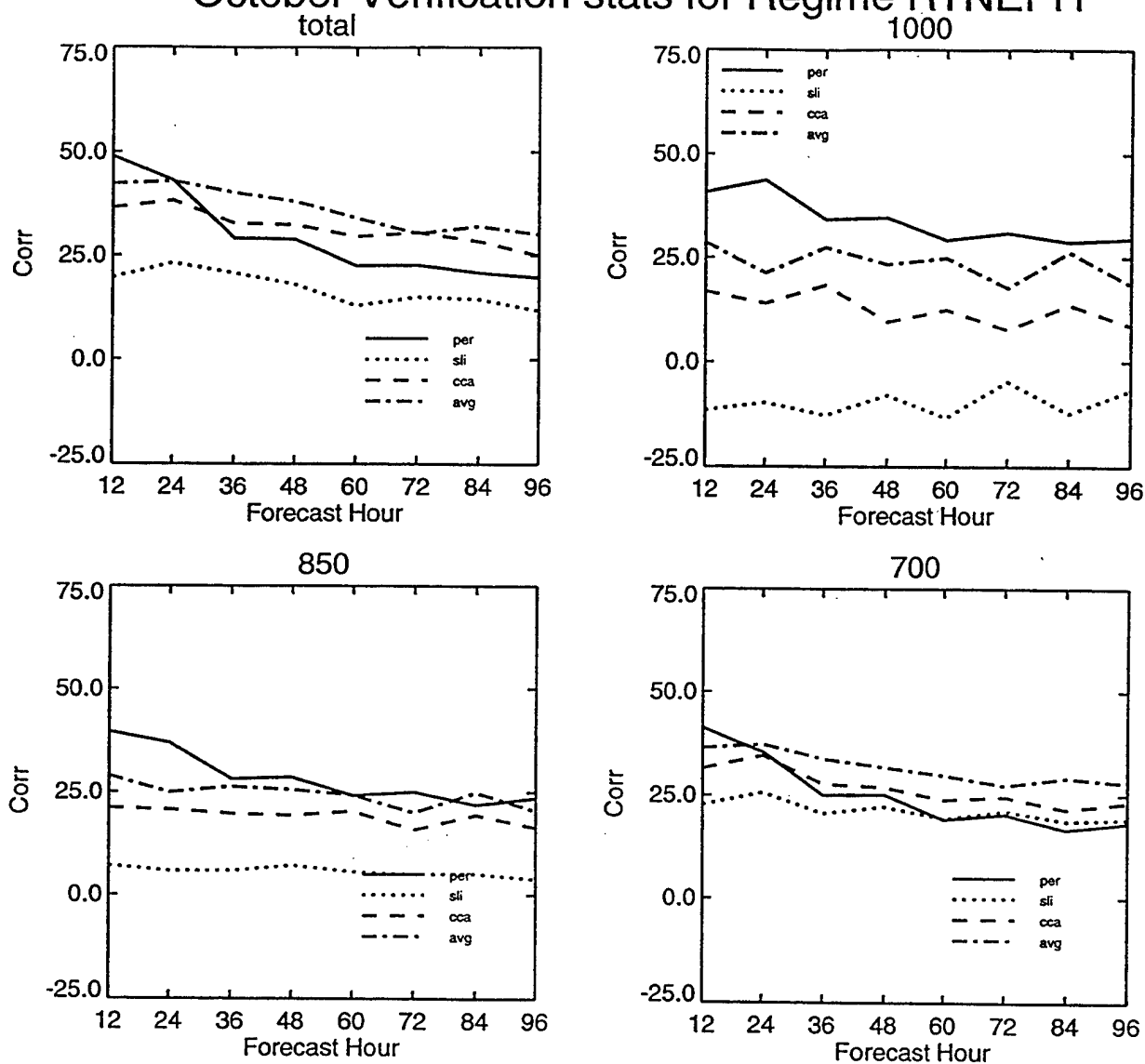


Figure 78: Same as Figure 58, except for the correlation of 12-96 hour forecasts of cloud cover for October, computed over the NH RTNEPH octagon.

# October Verification stats for Regime RTNEPH

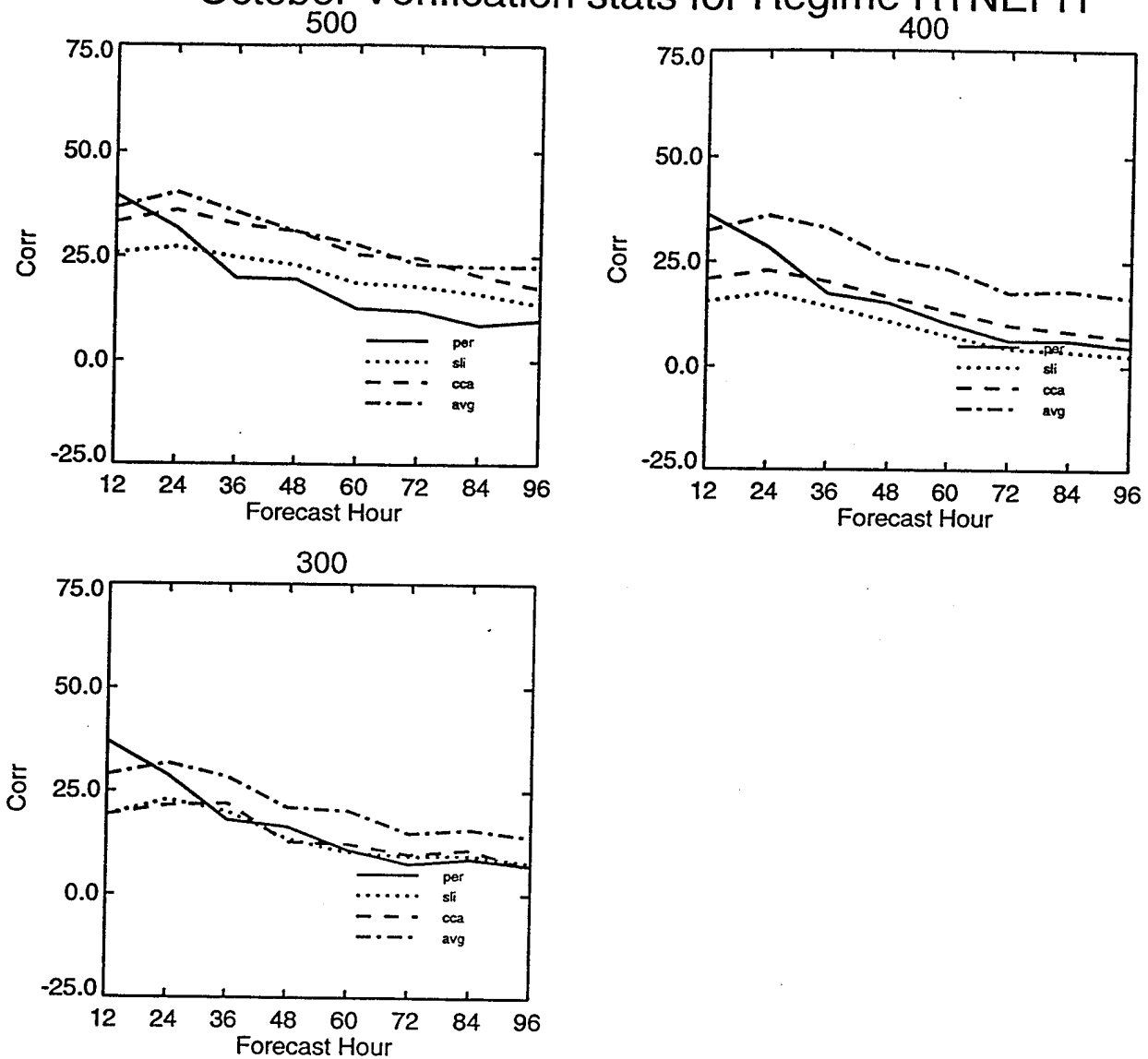


Figure 78: Continued.

# April Verification stats for Regime RTNEPH

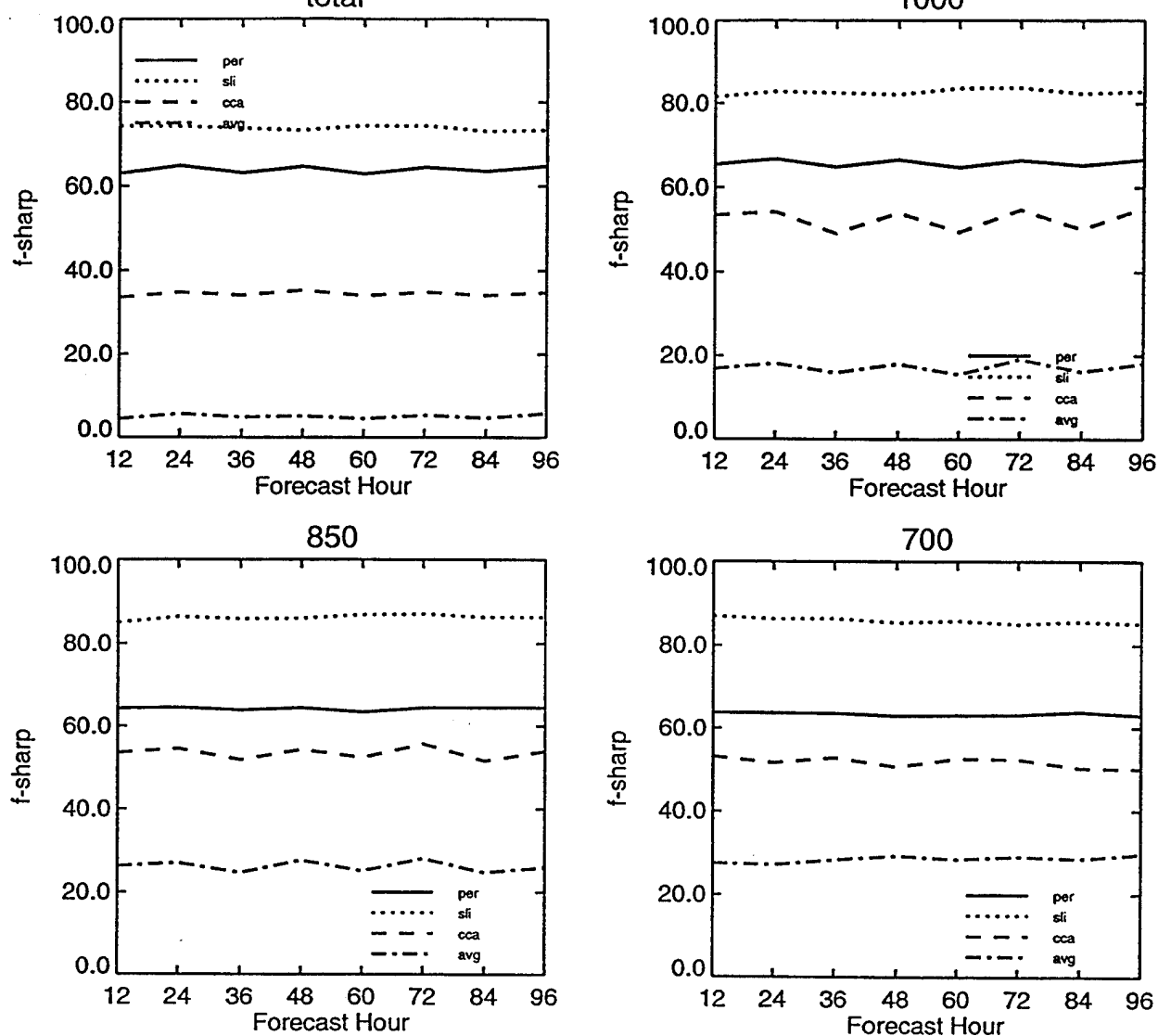


Figure 79: Same as Figure 58, except for the sharpness of 12-96 hour forecasts of cloud cover for April, computed over the NH RTNEPH octagon.

# April Verification stats for Regime RTNEPH

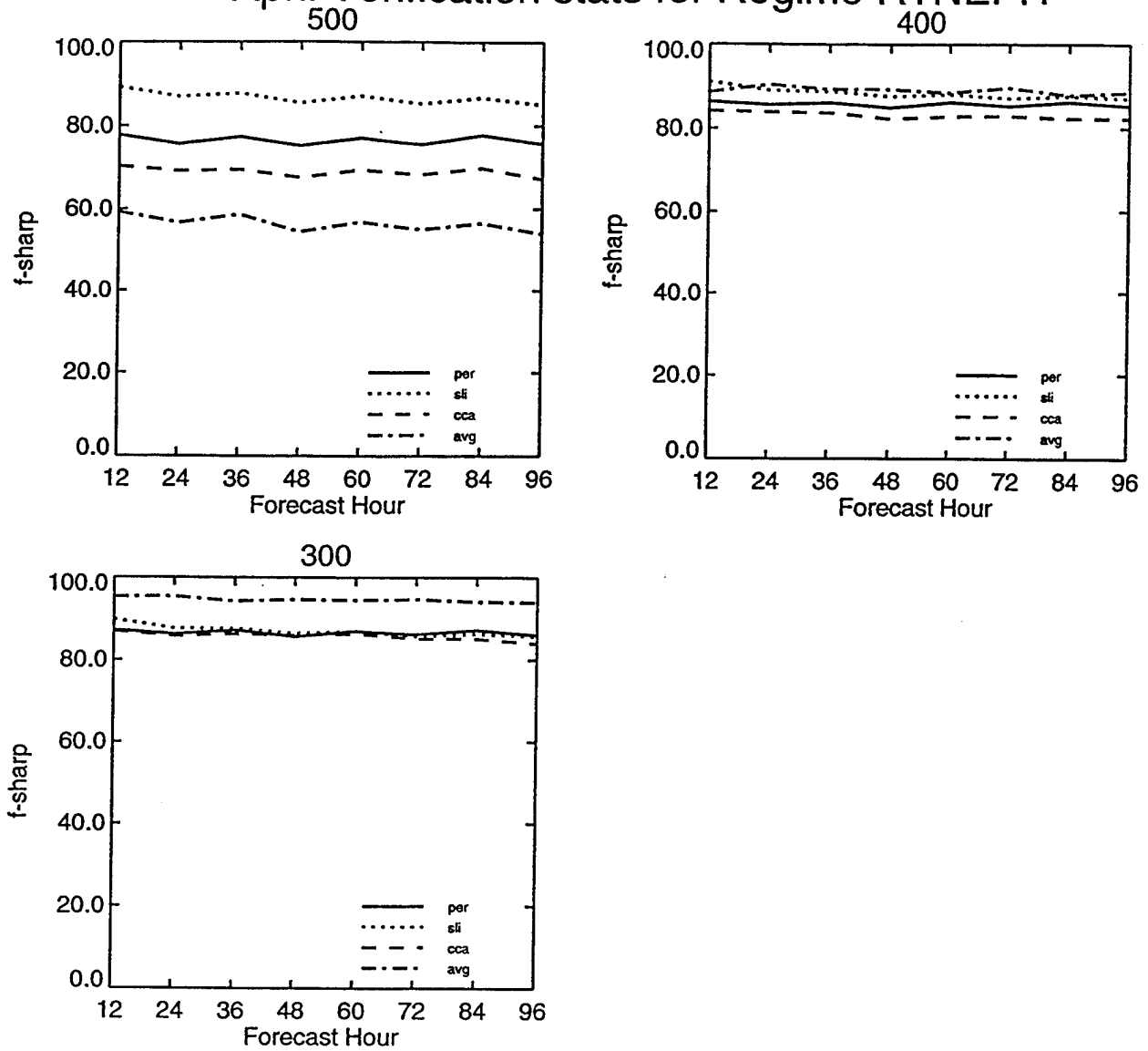


Figure 79: Continued.

## July Verification stats for Regime RTNEPH

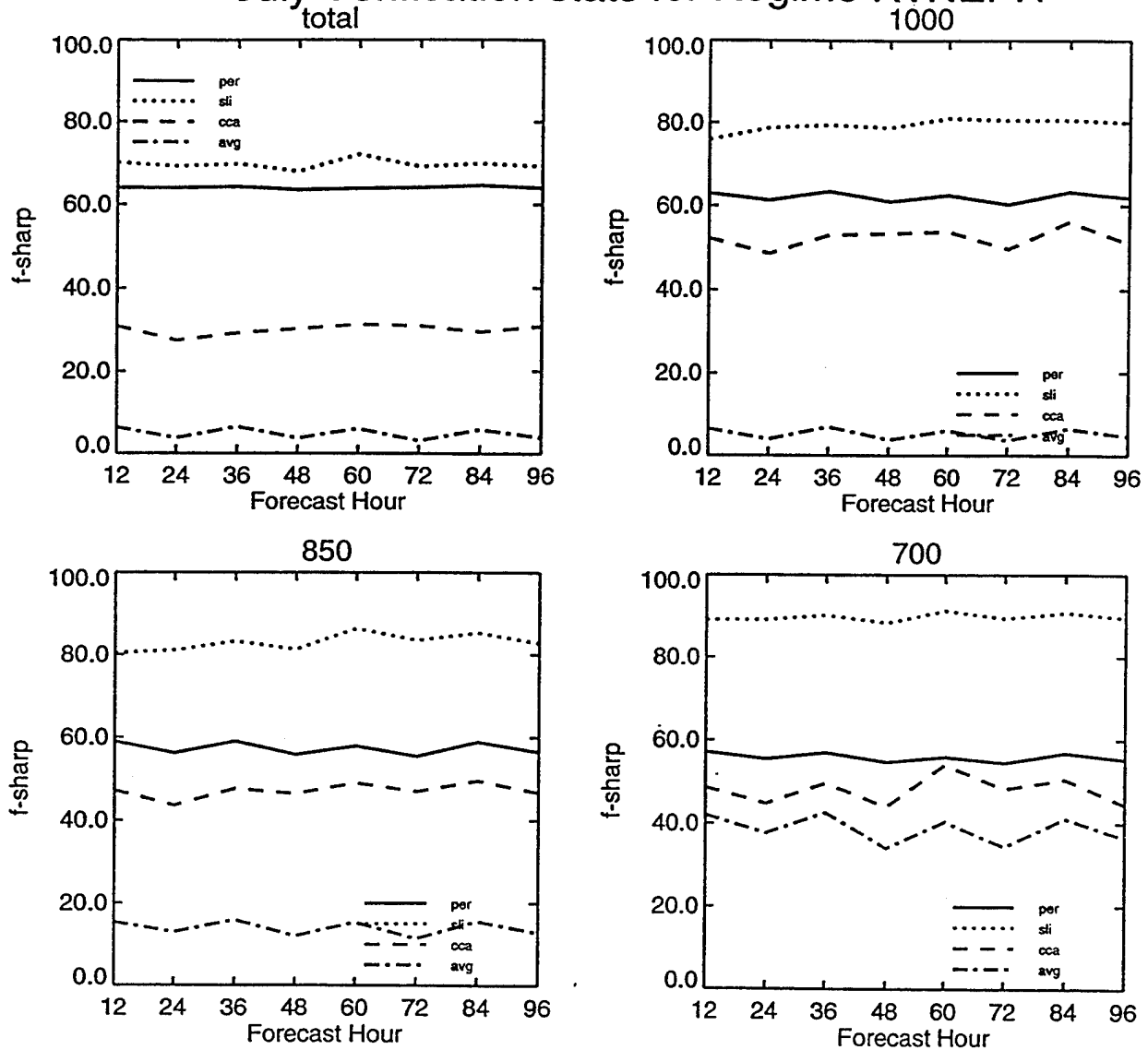


Figure 80: Same as Figure 58, except for the sharpness of 12-96 hour forecasts of cloud cover for July, computed over the NH RTNEPH octagon.

# July Verification stats for Regime RTNEPH

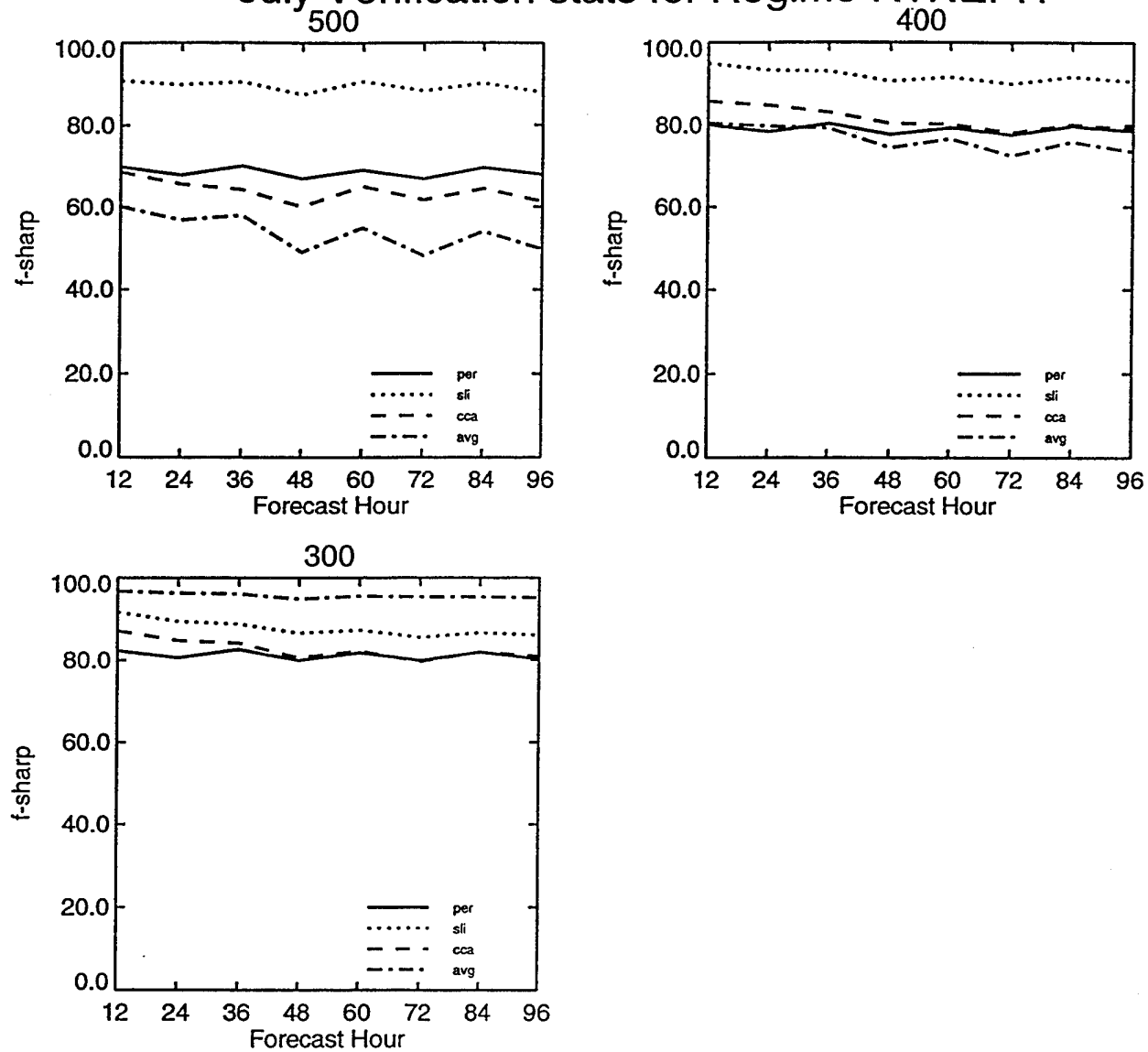


Figure 80: Continued.

## October Verification stats for Regime RTNEPH

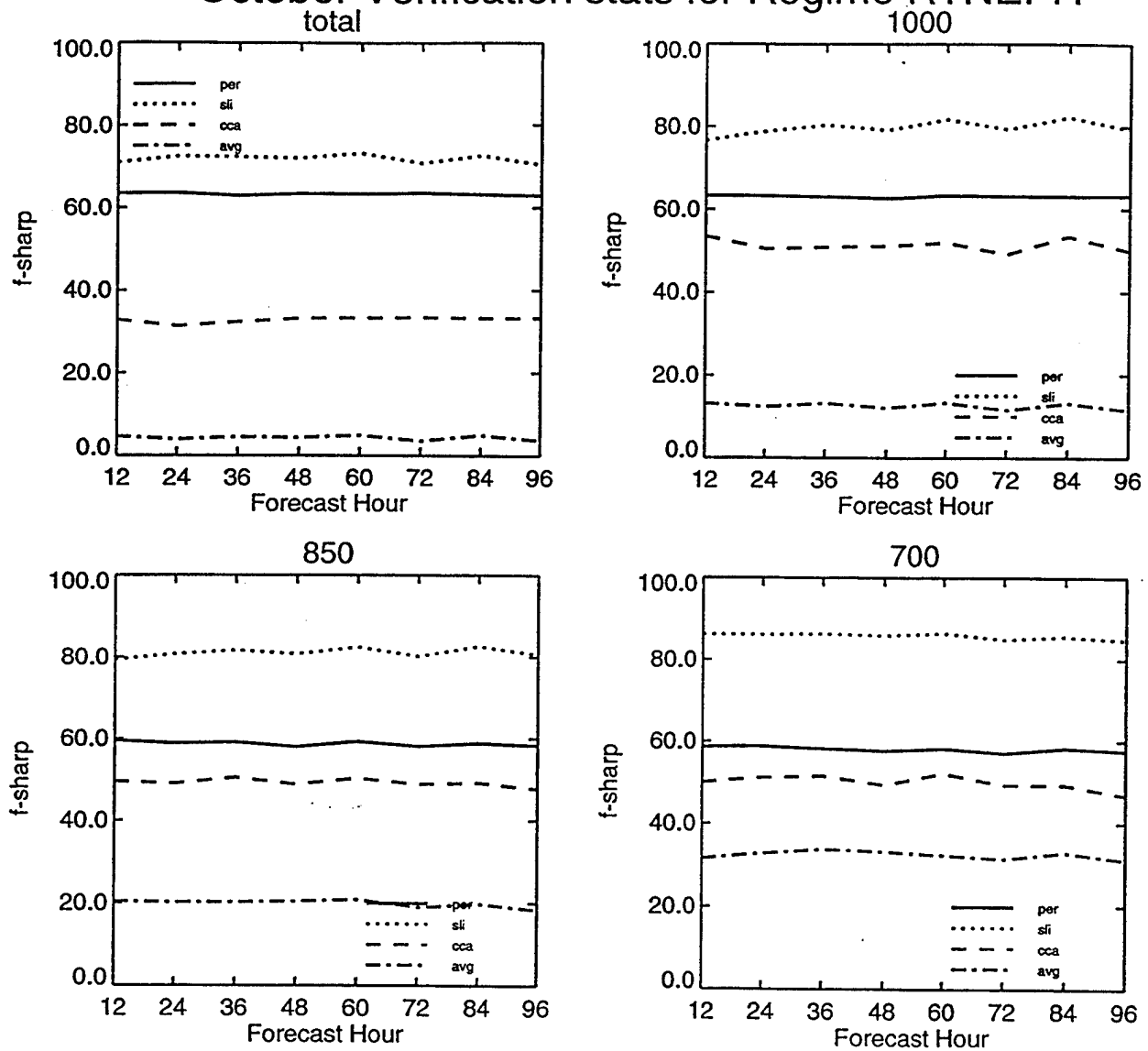


Figure 81: Same as Figure 58, except for the sharpness of 12-96 hour forecasts of cloud cover for October, computed over the NH RTNEPH octagon.



# October Verification stats for Regime RTNEPH

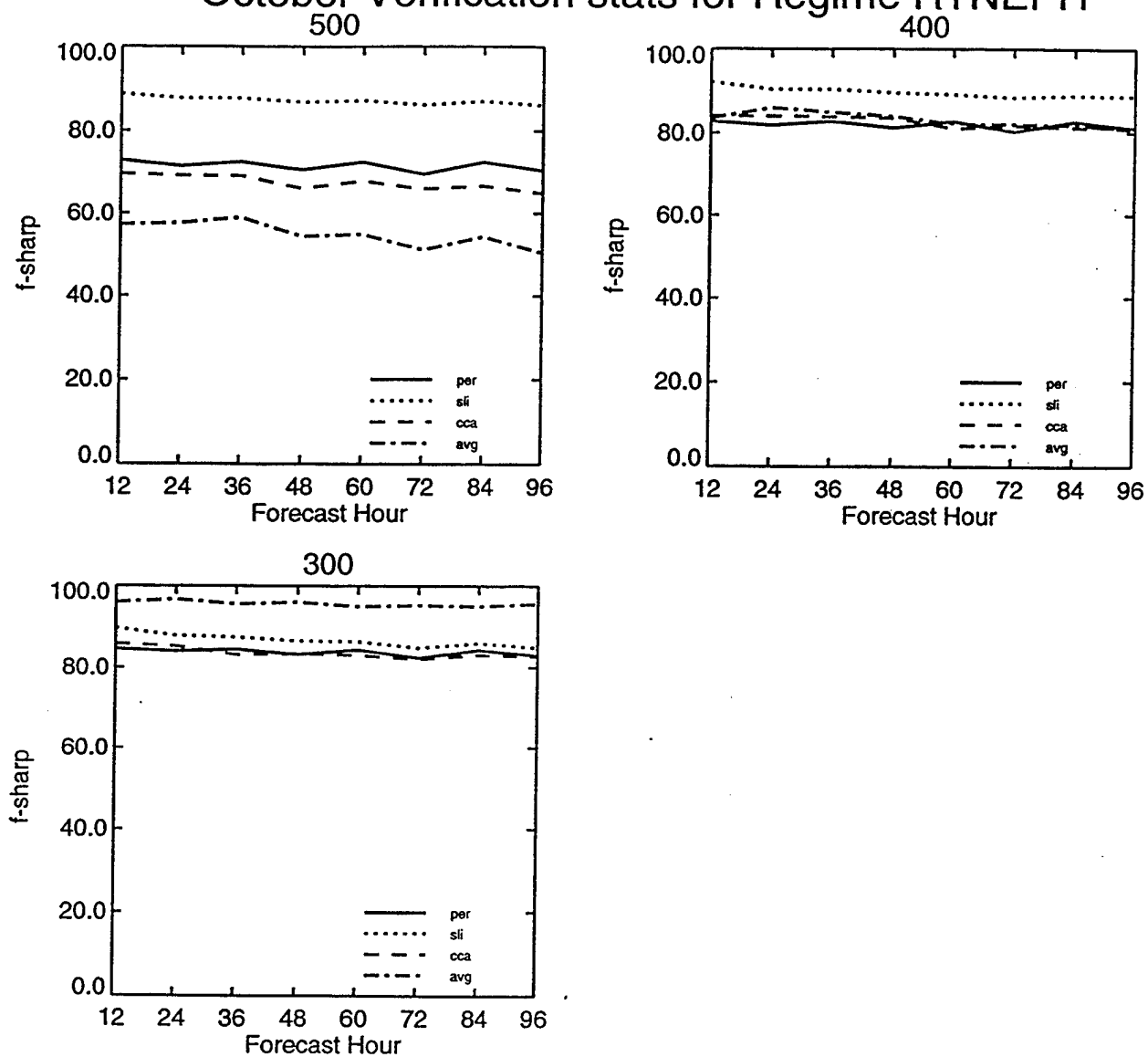


Figure 81: Continued.

**Appendix A: Manuscript accepted for publication in the International  
Journal of Climatology**

Hierarchical Clustering of Atmospheric Soundings

Marina Živković

Atmospheric and Environmental Research, Inc.

Cambridge, MA 02139

USA

## ABSTRACT

This is a procedural paper which compares six commonly used hierarchical clustering techniques applied to atmospheric soundings. The following techniques are compared: single-, complete- and two average-linkage techniques, the Ward's technique and one centroid type technique. Vertical profiling of the atmosphere is based on common soundings sampled on synoptic scales over the Northern Hemisphere. At each analysis point, the atmospheric profile is represented by an "atmospheric state vector" consisting of surface pressure, total column (precipitable) water, and temperature, wind and relative humidity values at twelve discrete vertical levels.

An evaluation of the techniques is performed by comparing the mean state vectors of final clusters. Four techniques produce comparable results with the largest differences between the techniques appearing for the surface pressure component of the cluster state vector. Also, large differences are found in the temperature and wind vector components at the 850 hPa pressure level, and the smallest differences are found for the relative humidity components. The results support findings of studies on air-mass typing on local or regional scales: the average-linkage method based on group average provides the most distinct and homogeneous clusters.

**KEYWORDS:** Hierarchical clustering, principal components analysis, weather typing, atmospheric soundings, synoptic climatology.

## INTRODUCTION

Hierarchical clustering procedures are used to objectively classify local weather conditions. Weather typing and synoptic weather regimes are usually defined by a similarity of synoptic maps and are used extensively in diagnostic studies (e.g., Klein, 1965; Peagle and Kieruff, 1974; Brazel and Nickling, 1986). The objective classification of weather regimes in synoptic climatology studies consists of first reducing the climatic data set by some type of factor analysis (e.g., Puvaneswaran, 1990). The weather categorization is then obtained by applying a clustering procedure (e.g., LeDrew, 1985; Wolter, 1987; Huth *et al.*, 1993).

The typified weather conditions can then be related to other variables of interest in environmental applications. For example, Kalkstein and Corrigan (1986) used synoptic weather typing to relate sulfur dioxide concentrations to individual weather variables, and Fernau and Samson (1990) used a similar approach to relate precipitation chemistry to air mass transport types. The temporal characteristics of weather types are used to define regional seasons (Alsop, 1989), climate (Fovell and Fovell, 1993;), climate change (Kalkstein *et al.*, 1990) or the impact of individual atmospheric elements to climate change (Hay *et al.*, 1992).

One possible approach to objective weather typing is a categorization based on prevailing air-mass over the observational site (Davis and Kalkstein, 1990). Similar weather conditions are grouped into weather types by the hierarchical clustering procedure, comprising a temporal synoptic index. The index is related to a prevailing air-mass and daily changes according to synoptic circulation. This approach is often based on surface observations and/or upper-air data at selected levels (e.g., Davis and Walker, 1992). Some applications, however, can benefit from

considering more detailed vertical soundings since the values of a single variable at different altitudes are not necessarily highly correlated under local conditions. Typing of such conditions can be used in applications that involve wider availability of global data sets. Živković and Louis (1992), for example, utilized global analyses to derive a cloud scheme applicable to general circulation models GCM. Objective typing of the soundings proved to be beneficial for specifying cloud cover on a GCM resolution scale. Garand (1993), in a reverse approach, typed cloud patterns to derive a procedure for relative humidity retrievals from satellite observations. These kinds of studies involve information on the entire vertical air-mass composition and on hemispheric or global scales which have not yet been addressed in synoptic climatology studies.

Objective classification can be very sensitive to the choice of hierarchical clustering method. Kalkstein *et. al.* (1987) showed that the average linkage method can be superior to Ward's technique for definition of a temporal synoptic index, whereas Fernau and Samson (1990) found the Ward's technique to be more applicable to the study of precipitation chemistry. The present study examines this sensitivity as applied to atmospheric vertical soundings on large spatial scales. Six hierarchical clustering techniques are examined for the classification of temperature, wind speed and relative humidity soundings.

## PROCEDURE

### *Data*

The comparison of the clustering techniques is prompted by their potential applications in general circulation models. For that reason vertical soundings derived from a sample of operational analyses from the US Air Force Global Weather Central (GWC) are used, with a sampling procedure described in Živković and Louis (1992).

The analyses provide data on a regular global grid with approximately 100 km resolution and at the first 12  $\sigma$ -levels used for integration of the Phillips Laboratory Global Spectral Model (Norquist *et al.*, 1992). Because this large number of grid points is redundant for the present study, only points spaced by synoptic scale (1000 km) are selected for each observational period. Also, the region is limited to the Northern Hemisphere only, producing a sample of 292 points for a single observational period (Fig. 1).

Analyses are available every twelve hours, which are again too frequent, and additional selection of observations in time is made. Since most of the synoptic processes are considered to be on a weekly time scale, analyses separated by one week are retained. Also, for the purpose of testing clustering algorithms it is sufficient to consider data for one month only, which provides a large enough sample. This selection procedure reduces a one month sample of complete analyses to a sample consisting of 4 observational periods for 292 points (total of 1168 profiles). All the profiles are derived from the analyses made at 0000 GMT during the month of January 1989.

At each grid point, vertical profiles of temperature (T), wind speed ( $|V|$ ) and relative humidity (RH) are available at 12 discrete levels ( $\sigma$  -

layers in the first column of Table 1). This yields 12 thermal variables, 12 dynamical variables and 12 moisture variables for each point. In addition to these 36 variables, the observations of two variables: surface pressure ( $p_s$ ) and precipitable water ( $w$ ) are also retained in the analysis. Hence, the atmospheric vertical profile at a single grid point is described by an 38-component vector

$$\mathbf{X} = (T_1, \dots, T_{12}, |V_1|, \dots, |V_{12}|, RH_1, \dots, RH_{12}, p_s, w). \quad (1)$$

This vector is called herein, an atmospheric state vector.

For the purpose of cluster analysis, our data sample is a matrix of 1168 realizations of a vector  $\mathbf{X}$ , *i.e.*,

$$\begin{Bmatrix} \mathbf{X}_1 \\ \mathbf{X}_j \\ \mathbf{X}_{1168} \end{Bmatrix} = \{x_{ij}\} \quad (2)$$

where each row of the matrix represents an independent sounding  $\mathbf{X}_j$ .

The procedure of identifying similar atmospheric states, and the associated weather regimes, is equivalent to identifying similar rows of matrix  $\{x_{ij}\}$ .

Matrix (2) is heterogeneous since it consists of variables that are differently scaled and correlated. A comparison of probability density functions for each variable illustrates this heterogeneity. An example is given in Fig. 2 where the probability distributions of five components of the state vector are shown. On the other hand, the vertical components of the variables, for example temperature, are correlated and have similar distributions.

For many climatological applications it is acceptable to analyze variables of heterogeneous scales, and it is their variance that determines weather regimes. To identify the typical modes of this variance, it is

necessary to reduce all the variables to comparable scales. This is accomplished by subtracting a sample mean from each state vector component and normalizing by its standard deviation. The mean values and standard deviation for each of the components presently used are listed in Table 1.

The values in Table 1 are also indicative of some general sample characteristics. For example, the sample mean temperature and dispersion (standard deviation) decrease with height. Mean wind speed dispersion increases with height. Finally, mean relative humidity decreases with height but its dispersion increases with height. The latter increase of the dispersion is not surprising given the uncertainties in observations and natural variability of the upper-level atmospheric moisture. Temperature and wind speed are more accurately analyzed variables, although their dispersions change with height at somewhat higher rates than the corresponding mean values.

#### *Principal components analysis*

The procedure of identifying similar weather regimes is equivalent to a procedure of identifying similar rows of matrix (2). Given its dimension and heterogeneity, this operation can be computationally extensive. Since the variance of a state vector  $\mathbf{X}_j$  typifies different regimes, it is possible to simplify this procedure by a principal components (PC) analysis of matrix (2). A common rule for retaining PCs is to retain the ones that resolve the variance above the noise level for a given process (*e.g.*, Overland and Preisendorfer, 1982) as described by their eigenvalues. In the present analysis, the linear independence of the PCs also removes the interdependence of the state vector components.



Figure 3 illustrates the fraction of variance resolved by the first ten principal components derived from the present sample. The first three components resolve over 78% of total variance, while the first ten components resolve 98.5% of variance. The corresponding PCs are plotted in Fig. 4. The horizontal axis in this figure represents indices for 38 variables from equation (1). The first twelve variables are the thermal variables. They are followed by twelve dynamic and twelve moisture variables. Finally, the surface pressure and precipitable water variables are marked by indices 37 and 38, respectively. The amount of variance resolved by each PC is given as a percentage value in the lower left-hand corner of each diagram. The first component resolves 43% of total variance, with the largest portion of it in thermal and dynamic variables (first 24 indices on the horizontal axis). The second component resolves most of the variance in the dynamic variables but resolves only 18% of total variance. Most of the variance resolved by the third component is in the moisture variables, etc. The picture becomes more complicated as the component number increases. Note that the variance explained by the surface pressure and precipitable water variables in the first five PCs is much less than the variance explained by the other 36 variables. This is indicative of a bias that the present PC analysis has towards the vertically sampled fields. The contribution of this bias is somewhat reduced by retaining ten components in the final analysis.

### *Cluster analysis*

Once the principal components are defined it is easy to project the original atmospheric state vector at a point,  $j$ , onto the PCs,  $\mathbf{b}_i$ ,

$$\mathbf{X}_j = \sum_{i=1}^n y_{ij} \mathbf{b}_i \quad (3)$$

where  $y_{ij}$  is the score for the  $i^{th}$  principal component at the point  $j$ . Similar atmospheric state vectors will have similar scores for the most dominant PCs. The goal of cluster analysis is to group rows of matrix  $\{y_{ij}\}$  that contain similar values for each retained PC (column values). First, it is necessary to define a measure of association among the scores that can be evaluated on a given sample. A common type of measure is a metric measure, and here Euclidean distance is used. Once the association between the scores is evaluated (i.e., a distance matrix is formed) it is necessary to define a method of grouping the points typified by similar scores. This is accomplished using hierarchical clustering.

### *Clustering Methods*

The evaluated distances between each pair of state vectors (transformed into PC's) are stored as a distance matrix. A clustering procedure essentially represents a process of grouping similar elements of the original or modified distance matrix. The goal is to form groups of the observational points with scores that differ as little as possible from each other. At the same time the groups should be as distinctive as possible.

Clustering procedures are fully objective methods based solely on the distance values and their modification by the merging process. There are many clustering procedures available in mathematical libraries that mostly differ by the type of merging process. The procedural differences are significant and can lead to substantially different results, therefore we evaluated six clustering methods available in the IMSL STAT/LIBRARY: a single linkage method, complete linkage method, average linkage based on distances within the merged cluster, average linkage based on distances between two clusters, Ward's method and a centroid method. The methods

have been described in detail by Adenberg (1973) and IMSL user manuals; only the main characteristics of the methods are given here.

*a) Single linkage method*

The criterion for merging clusters by this method is a distance determined by the minimal distance between two elements, one from each cluster. As the merger process continues, single entities are more likely to join an existing cluster than to merge with another element and form a new cluster. This is known as a chaining or snow-balling effect that produces a single large cluster. The applications of the method are limited since it tends to produce a single cluster instead of distinctive clusters. In the present application the method has been included as a reference only.

*b) Complete linkage method*

A complete linkage method is essentially the same as a single linkage method but the criterion for merging clusters is the maximum distance between two elements, one from each cluster. As a result, all elements within the new cluster will have distances smaller than the distance used for a merger. The method is called complete linkage because all the elements within the cluster are linked to each other at some maximum distance.

*c) Average linkage based on distances within a cluster*

In this method, an average distance is calculated between all the elements that would exist in a new merged cluster, including ones that have already been merged. A new cluster has the smallest average distance between all of its elements. The use of an average distance instead of an extreme distance reduces the distance dispersion within a cluster when compared to method a). This method is referred herein as an "average-in" method.

*d) Average linkage based on group distances*

This method is very similar to method c) but the average distance is calculated only from the distances that exist between the elements of two clusters considered for merging. The distances between elements that are already within the same cluster are not included in the average. This is often referred to as a "group average". Clusters with the smallest average distance between their groups of elements are merged. As shown by Kalkstein *et al.* (1987), this method produces the smallest distance dispersion within a cluster in some meteorological applications. It also maximizes the distances between the clusters, producing quite distinctive clusters. This method is referred herein as an "average-out" method.

*e) Ward's method*

This method is very widely used. The method merges clusters based on a criterion of maximizing an objective function separately defined for each particular application. In the present study, the merging process minimizes the total increase of cluster variance based on Euclidean distances. This procedure tends to produce uniformly sized clusters and is sample size biased. In some applications to spatially heterogeneous fields of a single variable (for example, precipitation) it can be superior to other methods in minimizing the cluster variance (Fernau and Samson, 1990). However, other methods such as rotated principal components may produce even more homogeneous classes in such cases (Gong and Richman, 1995)

*f) Centroid method*

The centroid method tested here is actually a modification of the centroid method available from the IMSL STAT/LIBRARY. The merging procedure is based on a calculation of the distance between the "centers" of

the clusters to be merged. Clusters with the least distant centers are merged. As in method e), cluster centers are simply the mean cluster vectors which are recalculated after every merger. The merging criterion, however, is similar to method a) since it is applied to the distances between cluster centers rather than individual cluster elements. Since no objective function is used as a merging criterion, the distances between clusters vary at each merging step. This method is also characterized by a snow-balling process, but still distinguishes more clusters than the method a).

Originally this method was designed for the purpose of clustering very large data sets when the storage requirement for the distance matrix is substantial. Distances are recalculated after every merger and the whole procedure is computationally fast. An advantage of this procedure is that it distinguishes well between very different weather regimes when a large number of clusters is retained. In the application by Živković and Louis (1992) to global atmospheric fields, this method produced a large tropical cluster, but also a few quite distinctive clusters. The snow-balling effect was treated by retaining a large number of clusters (thirty six) in the initial clustering procedure. The large subtropical cluster was then clustered again (nested clustering) creating quite distinctive weather regimes.

#### *Comparison of Clustering Methods*

Although the clustering procedures are objective in the sense that they are based on a computed measure of similarity, there are subjective factors involved in creating the final clusters. One factor is to decide how many clusters should be retained in the final analysis and what is the threshold aggregation level for each of the procedures. This decision can be based on the total amount of variance resolved by the clusters after each merger. Kalkstein *et al.* (1987) utilized a  $R^2$  multiple correlation for this

purpose. They observed that the number of clusters vary from method to method. By applying the same  $R^2$  criterion to their sample of 620 elements, they obtained 9 clusters for the average-linkage method, 6 clusters for the Ward's method and 13 clusters for the centroid method. Huth *et al.* (1993) found that the  $R^2$  criterion for the same average-linkage method suggested using 32 as a threshold aggregation level when applied to the variance of individual variables rather than the combined variance. The problem with retaining such a large number of clusters is still the snow-balling effect, which creates an uneven cluster size distribution and may require nested clustering in some applications (e.g., Živković and Louis, 1992; Kalkstein *et al.*, 1990).

Since each of the methods uses different merging procedures, the clusters will not only differ in size but also in typical properties which will be reflected in the cluster variance. For applications where physical interpretation of the clusters is essential, the goal of the procedure is to produce clusters that consist of elements that are as similar as possible. A way of measuring this similarity is to calculate the variance within the cluster. As long as the clusters remain uniform their variance will change very slowly with each merger. Once a less uniform cluster is formed by an inappropriate merger, the cluster variance will change significantly. As described by Kalkstein *et al.* (1987) the merging process should be terminated just before the cluster variance shows a significant change in value.

Many studies indicate that the criterion for terminating a clustering procedure is not only subjective but can also depend on the particular application (e.g., Yarnal and White, 1987). Aggregation levels based on  $R^2$  scores can vary from method to method or from one variable to another. A

variable aggregation level (Huth *et al.*, 1993) or nested clustering make direct comparison of the methods difficult.

To avoid these concerns, we start with the assumption that in the present data set similar weather distributions should be obtained by different methods independently of the number of retained clusters. A simple evaluation of the cluster variance for each of the methods indicates that the optimal aggregation level varies between 9 and 16 clusters. For this reason, a fixed aggregation level is chosen with a number of clusters larger than would be optimal for any of the methods. Twenty-four clusters are selected to be retained by each of the procedures to avoid the possibility of inadequate mergers. Also, given the large data size, some similarity between the clusters still can be expected since a few mergers are needed for each procedure to reach the optimal aggregation level.

However, for the centroid method this number of retained clusters is still too small because of the pronounced snow-balling effect. As we will see, an aggregation level of 24 creates uniformly sized clusters at least for the methods that produce comparable clusters, and eliminates clusters generated by the centroid and the single-linkage methods from further comparison.

## RESULTS

Each of the methods is characterized by a different merging procedure, hence each method produces clusters of different type and size. Figure 5 illustrates differences in cluster size for the six methods. The twenty four clusters are ordered according to their increasing size.

It is noticeable that complete linkage, average-in and Ward's method tend to produce clusters of more or less uniform size. The single linkage method produces many clusters with only a few elements and one large

cluster, while the centroid method also produces one very large cluster but also four additional clusters that have more than 30 elements. The average-out method tends to produce clusters of all sizes, and falls between the other two method groups.

### *Cluster Mean Vectors*

Clusters are compared by calculating the mean vectors and the standard deviation for each of their components, as well as the mean vectors of their PC scores. In Table 2 they are sorted by decreasing mean temperature in the first model layer, *i.e.*  $T_1$ . Thus the cluster labeled "1" is the cluster warmest near the surface as produced by each method. Similarly, cluster labeled "24" is the cluster coldest near the surface. Values in parenthesis show the number of elements within each cluster.

A few observations can be made by comparing near-surface cluster temperature among different methods. As described earlier, the soundings are sampled on a nearly equidistant grid with 1000 km spacing. A simple geometrical consideration of Fig. 1 shows that half of the soundings are located in the tropical region where weather and associated atmospheric structure vary little compared to the weather of the middle latitudes. Intuitively, one can expect the cluster analysis to produce fewer tropical clusters, which are of a large size, than in the midlatitudes where the weather is more variable and a smaller number of soundings is available. Similarly, in the polar and subpolar regions the number of soundings is the smallest. Like the tropical weather, the polar/subpolar weather varies little, particularly during a single winter month. Hence, one can expect fewer and smaller polar/subpolar clusters.

To simplify our discussion, let us choose the sample mean temperature near the surface (282 K) as a reference temperature. Clusters



that have temperature equal or larger than the reference temperature are underlined in Table 2. It is obvious that the clusters produced by the single-linkage method do not meet any of the above expectations. The centroid method, which also suffers from the snow-balling does not produce much better clusters. The warmest cluster has the largest size, whereas 3 colder clusters of a significant size are also produced. Among the remaining four methods, the Ward's method equally produces warmer and colder clusters of relatively uniform size. This demonstrates a bias that this method has towards the uniform cluster size. The remaining three methods clearly produce less warmer than colder clusters, with the average-out method generating the coldest clusters of the smallest size and the warmest cluster of the largest size.

The cluster ordering given in Table 2 is also used to compare the cluster mean vectors for each of the methods. Only the thermal variables of the vectors are presented here, and are plotted in Fig. 6. The figure is interpreted as follows. Along the horizontal axis, clusters are ordered by decreasing near-surface temperature, hence the distribution of their mean temperature should resemble the meridional distribution of climatic regions, i.e., clusters no. 1, 2 etc. are tropical clusters and are followed by the subtropical and middle-latitude clusters. The last clusters are the coldest, sub-polar clusters. In the vertical direction, temperature decreases with height, so the upper-most curve in each of the six diagrams represents the first layer or near surface temperature used for the cluster ordering. The lowest curve in each of the diagrams represents the corresponding temperature in the upper-most model layer (layer 12).

A comparison of the temperature mean vectors shows that the temperature varies smoothly with height for the methods that produce

clusters of a reasonable size. The single linkage and centroid methods show most of the variations in vertical structure from cluster to cluster because many of their clusters consist of very few vertical soundings. For larger clusters, for example clusters 13 and 14 obtained by the centroid method, variation is smoothed by the averaging.

An interesting feature is cluster numbered 22 obtained by the average-out method. This cluster is typified by stable lower troposphere (no-change or increase in temperature with height) in Fig. 7, moderate wind speed uniform with height and increased upper-level moisture (around 850 hPa - level 5). This cluster is not captured by the other comparable methods. (Note that all comparable methods would have a few more mergers before the "optimal" cluster number based on  $R^2$  scores would be reached). The single linkage and the centroid methods distinguishes a single profile corresponding to this weather situation, but among the remaining four methods only the average-out method retained this feature as a separate cluster. Note that this cluster contains only eight elements, but with a sample that contains more mid-latitude soundings, it could be identified as a larger cluster. The average-out method thus has a capability of identifying such distinctive structures even when a smaller number of clusters is retained. Other methods that produce clusters of more uniform size have assimilated this cluster at an earlier merger. This is an example where the clustering aggregation level depend on the cluster analysis application.

### *Within- and Between- Cluster Standard Deviations*

The four methods that produce comparable results are also compared by calculating within-cluster and between-cluster standard deviation at each vertical level as in Kalkstein *et al.* (1987). Within-cluster standard deviation is a measure of cluster homogeneity, while between-cluster standard deviation is a measure of cluster separation. Optimally, the method that produces the most distinct and the most homogeneous clusters would produce the largest between-cluster standard deviation and the smallest within-cluster standard deviation, respectively. The formulae used by Kalkstein *et al.* (1987) have been applied here.

Since the between- and within- cluster variances are related (as described by appropriate standard deviation), the ratio of the two types of standard deviations is calculated. These values are listed in Table 3 for each of the state vector components. The methods that produce the most distinguishable and homogeneous clusters are typified by the largest values of variance ratio.

Clearly, among the methods, the average-out method (column 4) produces the highest values of the variance ratio for all of the state vector components. This is particularly noticeable for the thermal variables. The differences between the methods are largest for the  $T_5$  variable (and below) which roughly corresponds to the temperature at 850 hPa pressure level. The differences are somewhat decreased for the dynamic variables, and the smallest for the moisture variable. Among the dynamic variables, differences are largest at and below the 850 hPa level, while for the moisture variables the methods differ the least at the upper levels.

Interestingly, the remaining two state vector components, surface pressure and precipitable water, show different sensitivity to the choice of the clustering method. Among all of the components, the surface pressure is the most sensitive to the type of clustering method and is the best classified by the average-out method. Conversely, the other moisture variable, total column (precipitable) water is the least sensitive to the typing technique (see also Fig. 8).

## CONCLUSIONS

Hierarchical clustering procedures have been tested on hemispheric atmospheric soundings sampled on synoptic spatial and temporal scales. Following PC analysis on the original 38 variables, the first three principal components resolve 78.2% of the sample variance, while the ten first components resolve 98.5 % of the sample variance. The first two components have the largest loadings in the temperature and wind speed fields while the third component is dominated by the loadings in the humidity field. The humidity field loadings also dominate the other higher components. The cluster analysis is applied to the scores for ten principal components.

Six hierarchical clustering methods are tested with the same aggregation level (24 clusters). Sizewise, three of the methods are found to produce clusters of relatively uniform size and two other methods are dominated by snow-balling effects. One method, average-linkage based on group averages, produces clusters of all sizes. In terms of their mean state vectors, four out of six methods produce clusters/regimes that are comparable: complete-linkage, two average-linkage methods and the Ward's method.

The homogeneity of the clusters is compared by calculating the average cluster standard deviation for each of the four comparable methods (also known as within-cluster standard deviation). The difference between the clusters/regimes identified by each of the four methods is compared by calculating between-cluster standard deviations. One method, average-linkage based on group averages, produce clusters of all sizes that are also the most distinct and homogeneous. This comparison supports conclusions of Kalkstein *et al.* (1987) based on surface variables.

The degree of homogeneity and distinction varies, however, among the variables, being the most noticeable for thermal variables and the least noticeable for moisture variables. Also, differences between the methods are least distinguishable for the moisture variables. This can be explained by the fact that the humidity is the most uncertain and noisiest atmospheric variable. Hence, the high variability can benefit in a statistical sense from uniformly sized clusters such as the ones produced by the Ward's or average-in technique. A similar observation of the Ward's technique performance was made for other moisture related data (precipitation) by Fernau and Samson (1990). Differences between the methods are the most distinguishable for thermal and dynamical variables, where the linkage method based on group averages appears to be superior to the other methods. This indicates that for dynamically based applications with global data, a careful examination of the clustering methods is needed because of substantial differences in circulation and weather between mid-latitude, tropical, and polar regions.

ACKNOWLEDGMENT: This work has been sponsored by the Air Force Material Command Contract F19628-92-C-0092. Special thanks to Thomas Nehrkorn for his support and data which he kindly provided. The author is

grateful to anonymous reviewers for constructive comments that improved the manuscript.

## REFERENCES

- Alsop, T. J., 1989: The natural Seasons of Western Oregon and Washington, *J. Climatol.*, **2**, 888-896.
- Anderberg, M. R., 1973: Cluster Analysis for Applications. Academic Press, New York, pp. 59.
- Brazel, A. J. and W. G. Nickling, 1986: The relationship of weather types to dust storm generation in Arizona (1965-1980). *J. Climatol.*, **6**, 255-275.
- Davis, R. E. and Kalkstein, L. S., 1990: Development of an automated spatial synoptic climatological classification. *Int. J. Climatol.*, **10**, 769-794.
- Davis, R. E. and Walker, D. R., 1992: An upper-air synoptic climatology of the Western United States. *J. Climate*, **5**, 1449-1467.
- Fernau, M. E. and Samson, P. J., 1990: Use of cluster analysis to define periods of similar meteorology and precipitation chemistry in Eastern North America. Part I: Transport patterns. *J. Appl. Meteor.*, **29**, 735-750.
- Fovell, R. G. and Fovell, M.-Y. C., 1993: Climate zones of the conterminous United States defined using cluster analysis. *J. Climate*, **6**, 2103-2135.
- Garand L., 1993: A pattern recognition technique for retrieving humidity profiles from Meteosat or GOES imagery. *J. Appl. Meteor.*, **32**, 1592-1607.
- Gong, X. and M. B. Richman, 1995: An examination of cluster analysis Methodology in Geophysical Research. Part I: Precipitation data. *J. Climate*, **8**, (in press).

- Hay, L. E., McCabe, G. J., Jr., Wolock, D. M. and Ayers, M. A., 1992: Use of weather types to disaggregate general circulation model predictions. *J. Geophys. Res.*, **97**, 2781-2790.
- Huth R., Nemešová, I. and Klimperová, N., 1993: Weather categorization based on the average linkage clustering technique: An application to European mid-latitudes. *Int. J. Climatol.*, **13**, 817-835.
- IMSL STAT/LIBRARY FORTRAN Subroutines for Statistical Analysis, 1987, Version 1, Houston, Texas, USA, Vol. 2, 721-742.
- Kalkstein, L. S. and Corrigan, P., 1986: A synoptic climatological approach for geographical analysis: assessment of sulfur dioxide concentrations, *Ann. Assoc. Am. Geogr.*, **76**, 381-395.
- Kalkstein, L. S., Tan, G. and Skindlov, J. A., 1987: An evaluation of three clustering procedures for use in synoptic climatological classification. *J. Climate Appl. Meteor.*, **26**, 717-730.
- Kalkstein, L. S., Dunne, P. C. and Vose, R. S., 1990: Detection of climatic change in the western North American Arctic using a synoptic climatological approach, *J. Climate*, **3**, 1153-1167.
- Klein, W. H., 1965: Synoptic climatological models for the United States. *Weatherwise*, **December**, 252-259.
- LeDrew, E. F., 1985: The dynamic climatology of the Beaufort to Laptev Sea sector of the polar basin for the winters of 1975 and 1976. *J. Climatol.*, **5**, 253-272.
- Norquist, D. C., C.-H. Yang, S. Chang and D. C. Hahn, 1992: Phillips laboratory global spectral numerical weather prediction model. PL-TR-92-2225. Phillips Laboratory, Hanscom AFB, MA., pp. 154. ADA267293



- Overland, J. E. and R. W. Preisendorfer, 1982: A significance test for principal component applied to a cyclone climatology. *Mon. Wea. Rev.*, **110**, 1-4.
- Pagle, J. N. and L. P. Kierluff, 1974: Synoptic climatology of 500 mb winter flow types, *J. Appl. Meteor.*, **13**, 205-220.
- Puvaneswaran, M., 1990: Climatic classification for Queensland using multivariate statistical techniques. *Int. J. Climatol.*, **10**, 591-608.
- Wolter, K., 1987: The southern oscillation in surface circulation and climate over the tropical Atlantic, Eastern Pacific, and Indian Oceans as captured by cluster analysis. *J. Climate. Appl. Meteor.*, **4**, 540-558.
- Yarnal, B. and White, D. A., 1987: Subjectivity in a computer-assisted synoptic climatology I: classification results. *J. Climatol.*, **7**, 119-128.
- Živković, M. and Louis, J.-F., 1992: A new method for developing cloud specification schemes in general circulation models. *Mon. Wea. Rev.*, **120**, 2928-2941.

*Table 1. Sample mean and standard deviation for thirty eight model variables.*

$\sigma$ layer	Temperature (°K)		Wind speed (ms <sup>-1</sup> )		Relative Humidity		Surface Pressure (Pa)		Precipitable Water (kg m <sup>-2</sup> )	
	mean	st.dev.	mean	st.dev.	mean	st.dev.	mean	st.dev.	mean	st.dev.
0.995	282.00	17.54	7.80	4.50	.73	.18	96974.6	7791.3	22.31	18.80
0.981	281.55	17.36	7.82	4.51	.72	.18	-	-	-	-
0.960	280.83	17.09	7.91	4.56	.70	.18	-	-	-	-
0.920	279.40	16.56	8.19	4.76	.67	.20	-	-	-	-
0.856	276.99	15.74	8.88	5.33	.61	.22	-	-	-	-
0.777	273.38	15.11	9.39	5.89	.55	.22	-	-	-	-
0.688	268.30	14.81	10.52	6.96	.48	.24	-	-	-	-
0.594	261.53	14.57	12.29	8.77	.44	.24	-	-	-	-
0.497	253.16	14.16	15.08	11.06	.41	.25	-	-	-	-
0.425	245.54	13.56	17.48	12.98	.40	.25	-	-	-	-
0.375	239.42	12.80	19.32	14.29	.39	.25	-	-	-	-
0.325	232.84	11.42	21.32	15.53	.39	.26	-	-	-	-

Table 2. Cluster mean temperature at the first level for six clustering procedures. Clusters are ordered by decreasing temperature value. Numbers in parenthesis show the cluster size.

cluster number	Method					
	single (1)	complete (2)	average-in (3)	average-out (4)	Ward's (5)	centroid (6)
1	284.4 (2)	297.5 (112)	295.8 (131)	297.5 (448)	299.4 (66)	295.2 (613)
2	282.7 (1)	296.9 (132)	295.33 (183)	289.0 (141)	298.7 (81)	283.2 (1)
3	<u>282.3 (1138)</u>	294.5 (22)	294.7 (71)	286.4 (2)	297.6 (115)	282.8 (33)
4	280.4 (2)	294.4 (286)	292.5 (146)	286.1 (97)	296.1 (95)	<u>282.7 (1)</u>
5	279.9 (1)	290.3 (55)	281.4 (80)	285.3 (22)	294.9 (49)	280.8 (5)
6	278.3 (1)	286.7 (44)	286.9 (16)	283.8 (25)	292.3 (77)	280.4 (2)
7	277.1 (1)	284.0 (13)	285.5 (28)	282.7 (1)	291.4 (86)	279.0 (205)
8	274.4 (1)	282.8 (24)	285.3 (16)	<u>282.0 (3)</u>	287.8 (30)	274.4 (1)
9	269.9 (1)	282.6 (48)	<u>284.2 (20)</u>	279.8 (4)	287.7 (18)	274.3 (2)
10	269.2 (2)	<u>282.2 (15)</u>	280.4 (2)	276.2 (10)	286.1 (16)	267.7 (1)
11	269.0 (1)	279.8 (8)	275.1 (20)	274.7 (48)	285.2 (44)	267.4 (1)
12	267.7 (1)	279.6 (11)	274.9 (63)	274.4 (1)	<u>284.1 (44)</u>	266.3 (8)
13	267.2 (1)	272.6 (18)	274.6 (21)	271.4 (76)	279.2 (5)	263.2 (38)
14	267.0 (1)	272.1 (55)	274.0 (42)	270.6 (5)	275.2 (52)	255.8 (242)
15	266.7 (4)	271.5 (14)	267.5 (6)	267.1 (57)	275.0 (58)	255.4 (1)
16	261.9 (1)	271.2 (23)	266.9 (26)	265.0 (10)	271.3 (31)	255.3 (4)
17	254.9 (1)	270.2 (31)	265.8 (12)	263.5 (35)	266.6 (36)	247.8 (1)
18	246.6 (1)	264.8 (30)	262.1 (19)	259.7 (6)	265.7 (41)	247.5 (2)
19	246.3 (1)	263.4 (15)	260.7 (20)	255.1 (21)	265.2 (14)	246.3 (1)
20	246.1 (1)	256.5 (97)	254.1 (86)	250.2 (143)	264.1 (46)	246.1 (1)
21	244.1 (2)	255.9 (16)	253.8 (18)	247.2 (2)	253.7 (35)	244.1 (2)
22	243.7 (1)	252.2 (36)	252.3 (19)	246.3 (8)	252.0 (58)	243.7 (1)
23	243.2 (1)	248.2 (16)	248.6 (43)	244.8 (2)	251.0 (28)	243.2 (1)
24	231.9 (1)	245.9 (47)	248.4 (30)	243.7 (1)	242.4 (43)	231.9 (1)

Table 3. Ratio of between-cluster and within-cluster standard deviations for thirty eight variables and four clustering methods.

Variable	Method			
	complete	average-in	average-out	Ward's
T <sub>1</sub>	2.28	2.48	3.77	2.86
T <sub>2</sub>	2.30	2.51	3.81	2.89
T <sub>3</sub>	2.33	2.53	3.88	2.93
T <sub>4</sub>	2.37	2.56	3.96	2.98
T <sub>5</sub>	2.36	2.54	3.96	2.93
T <sub>6</sub>	2.31	2.50	3.80	2.79
T <sub>7</sub>	2.28	2.51	3.73	2.71
T <sub>8</sub>	2.29	2.52	3.72	2.71
T <sub>9</sub>	2.35	2.53	3.68	2.75
T <sub>10</sub>	2.38	2.59	3.62	2.76
T <sub>11</sub>	2.32	2.59	3.54	2.73
T <sub>12</sub>	2.13	2.36	3.18	2.56
V  <sub>1</sub>	1.72	2.46	2.65	1.76
V  <sub>2</sub>	1.77	2.49	2.76	1.83
V  <sub>3</sub>	1.86	2.52	2.94	1.94
V  <sub>4</sub>	1.99	2.52	3.24	2.14
V  <sub>5</sub>	2.02	2.36	3.37	2.30
V  <sub>6</sub>	2.09	2.26	3.15	2.38
V  <sub>7</sub>	2.06	2.13	2.73	2.26
V  <sub>8</sub>	2.09	2.11	2.65	2.24
V  <sub>9</sub>	1.96	2.03	2.54	2.08
V  <sub>10</sub>	1.94	2.02	2.58	2.00
V  <sub>11</sub>	1.99	1.96	2.52	1.90
V  <sub>12</sub>	1.70	1.83	2.38	1.74
RH <sub>1</sub>	1.42	1.54	2.18	1.42
RH <sub>2</sub>	1.50	1.64	2.27	1.42
RH <sub>3</sub>	1.50	1.64	2.27	1.64
RH <sub>4</sub>	1.58	1.82	2.08	1.73
RH <sub>5</sub>	1.50	1.62	1.85	1.69
RH <sub>6</sub>	1.43	1.83	1.69	1.69
RH <sub>7</sub>	1.25	1.69	1.57	1.64
RH <sub>8</sub>	1.40	1.69	2.00	1.69
RH <sub>9</sub>	1.38	1.47	2.00	1.54
RH <sub>10</sub>	1.47	1.47	1.86	1.47
RH <sub>11</sub>	1.31	1.4	1.86	1.31
RH <sub>12</sub>	1.18	1.4	1.50	1.18
Ps	1.30	1.61	3.27	1.41
w	2.31	2.62	2.43	2.78

## LIST OF FIGURES

- Fig. 1 Map of the data points sampled with 1000 km resolution from a regular Global Spectral Model grid.
- Fig. 2 Probability distributions for temperature, wind speed, relative humidity in the first model layer and distributions for surface pressure and precipitable water .
- Fig. 3 Distribution of sample variance explained by the first ten PCs.
- Fig. 4. Loading scores for the first ten principal components.  
Percentage of resolved variance is given at the lower left corner of each diagram.
- Fig. 5 Number of observations in twenty four clusters obtained by six clustering methods. Each method is represented by different line pattern: "single" represents method a), "complete" represents method b), "average-in" represents method c), "average-out" represents method d), "Ward's" represents method e), and "centroid" represents method f). Clusters are ordered by increasing size.
- Fig. 6 Cluster mean temperature vector obtained by six clustering methods. The upper-most curve represents temperature mean value at the lowest level for each of the six methods. Clusters are ordered by the decreasing lowest level temperature: cluster 1 is tropical and cluster 24 is polar cluster.
- Fig. 7 Vertical temperature profiles that comprise cluster 22 obtained by the average-out method.
- Fig. 8 Cluster mean precipitable water and standard deviation obtained by six clustering methods. Solid line pattern represents cluster mean and dashed line pattern represents  $\pm$  standard deviation. Clusters

are ordered by the decreasing lowest level temperature: cluster 1 is tropical and cluster 24 is polar cluster.

Data Points

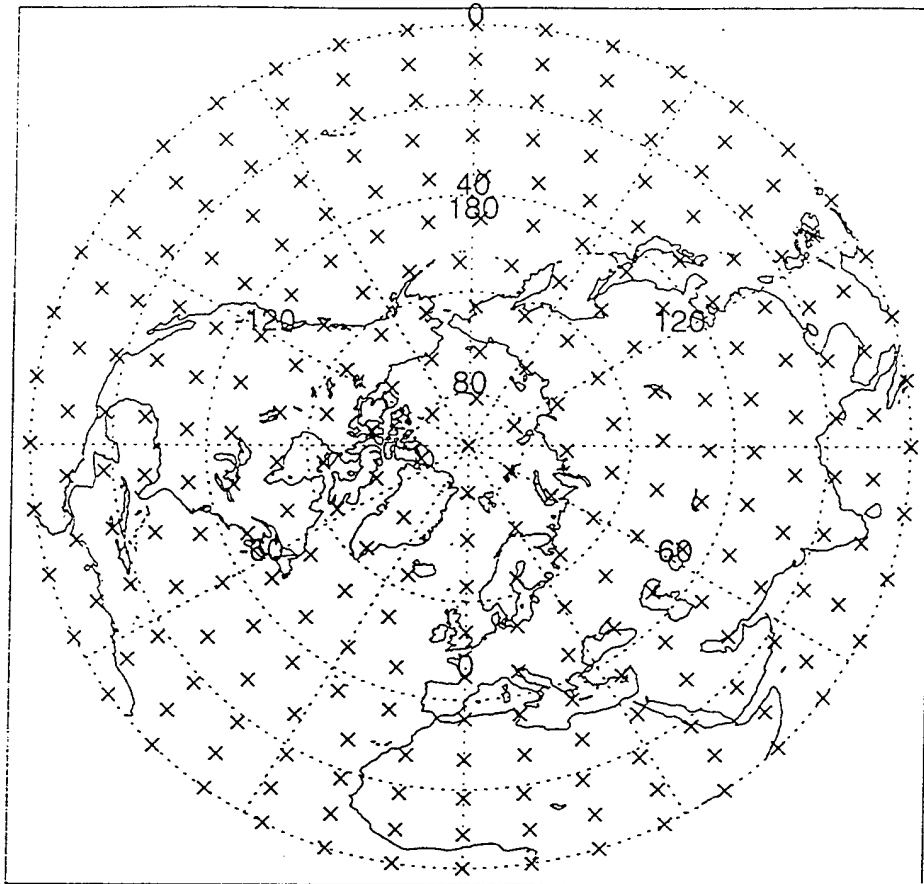


FIGURE 1

## Probability Distribution

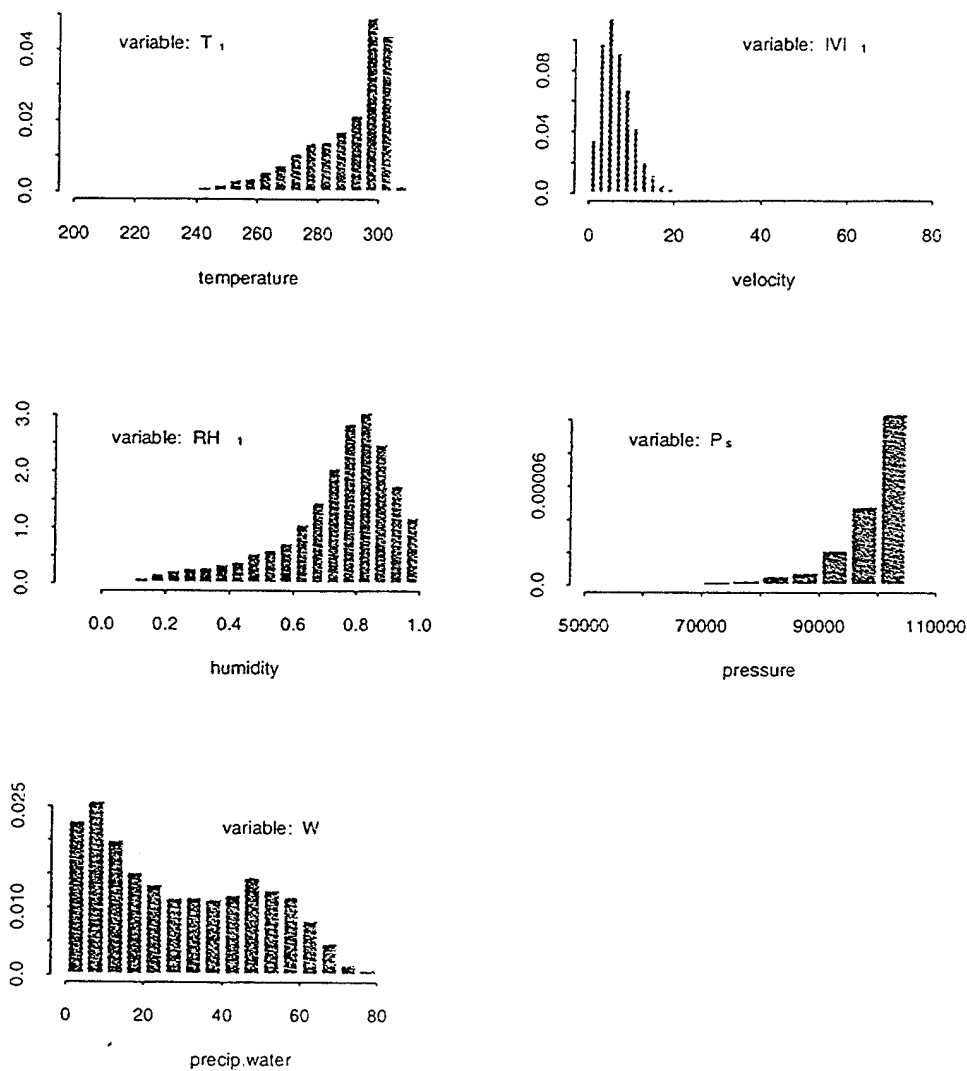


FIGURE 2



## Northern Hemisphere Data

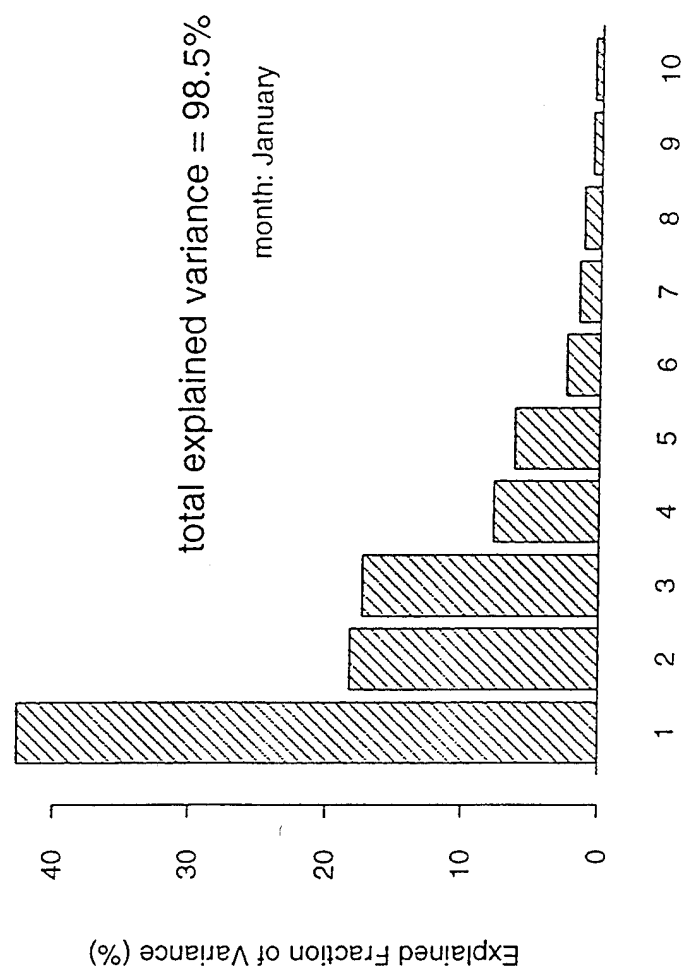
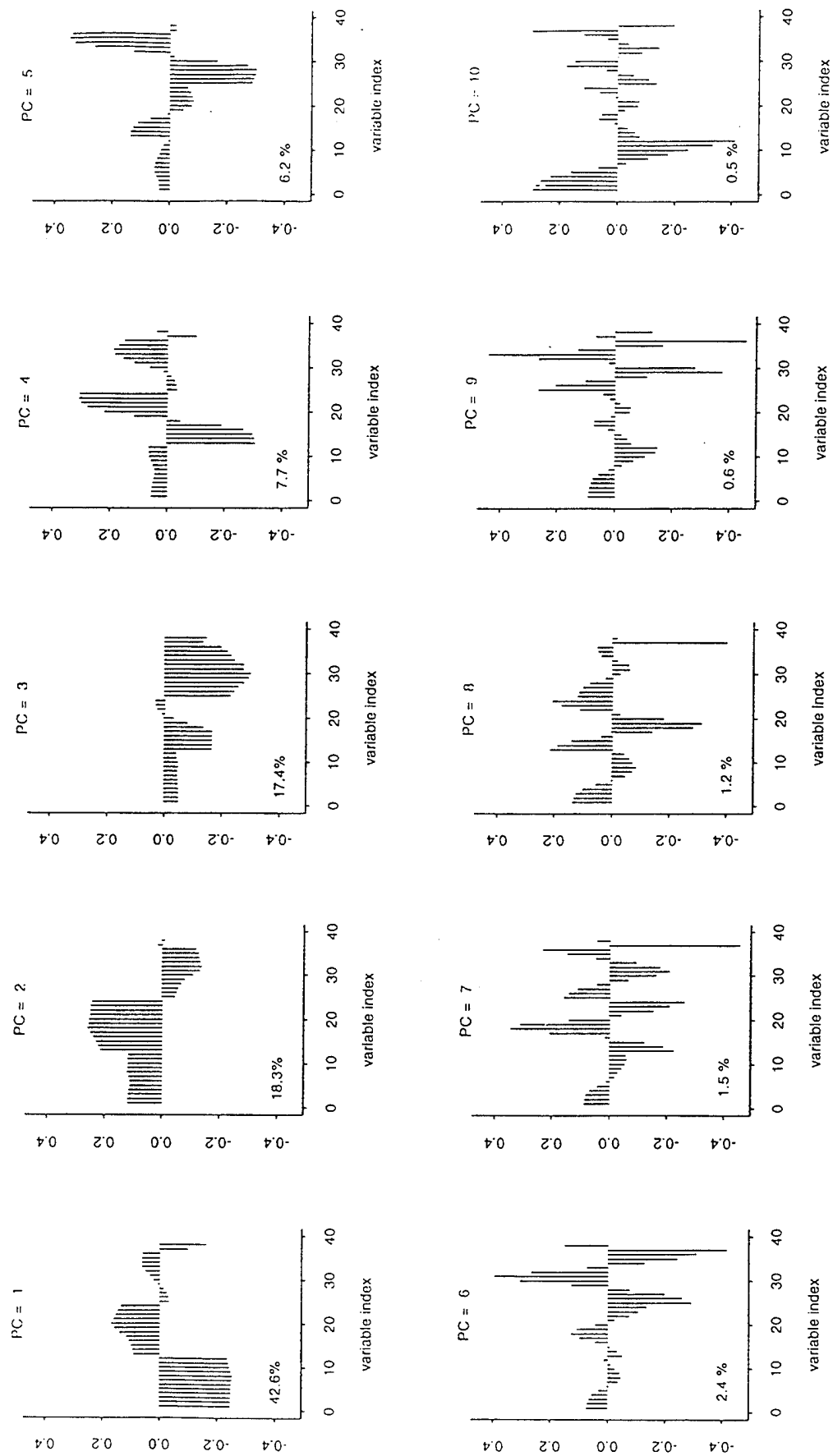


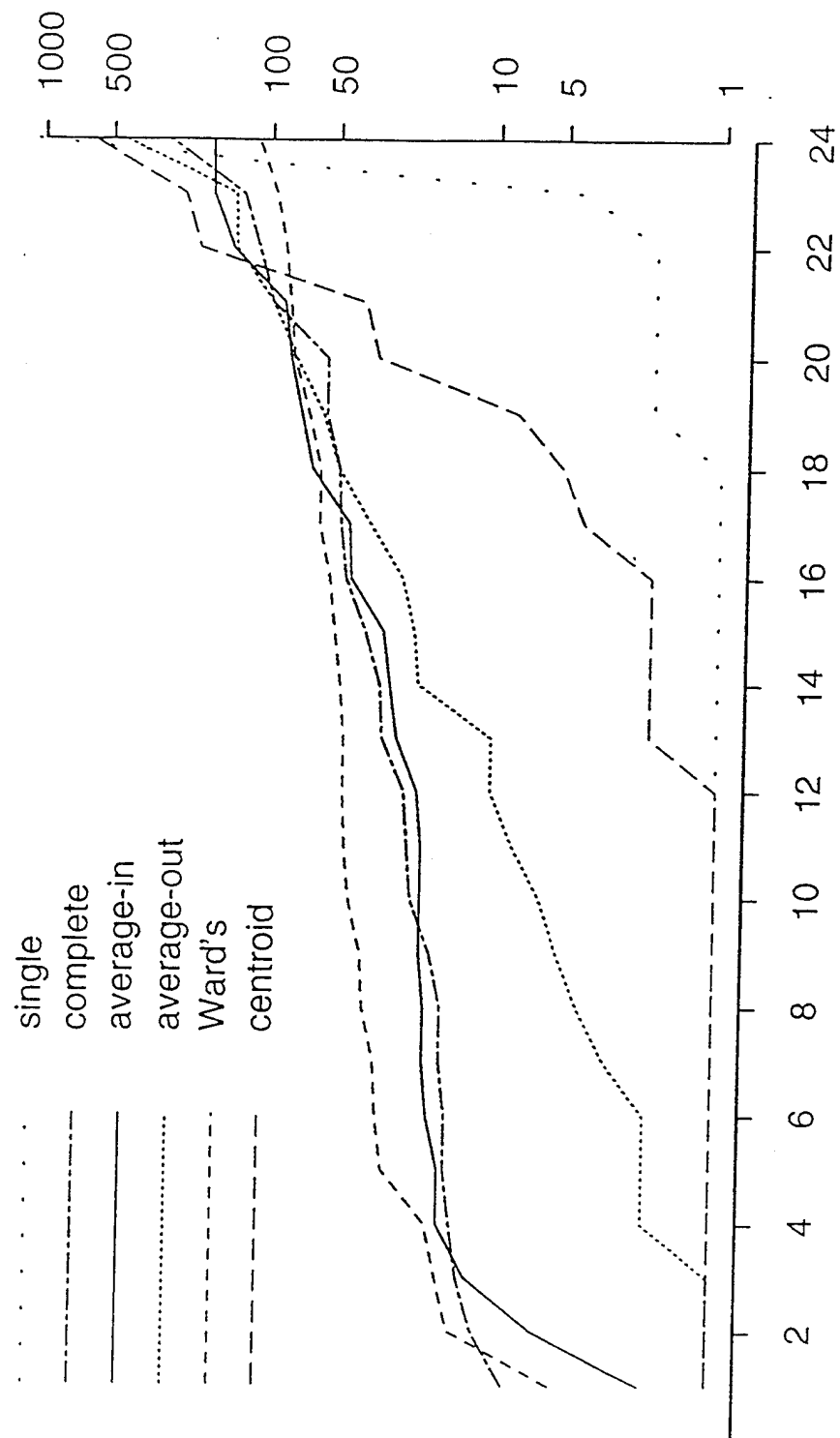
FIGURE 3

# First Ten Principal Components



**FIGURE 4**

# Cluster Size



cluster number

FIGURE 5

# Temperature Cluster Mean

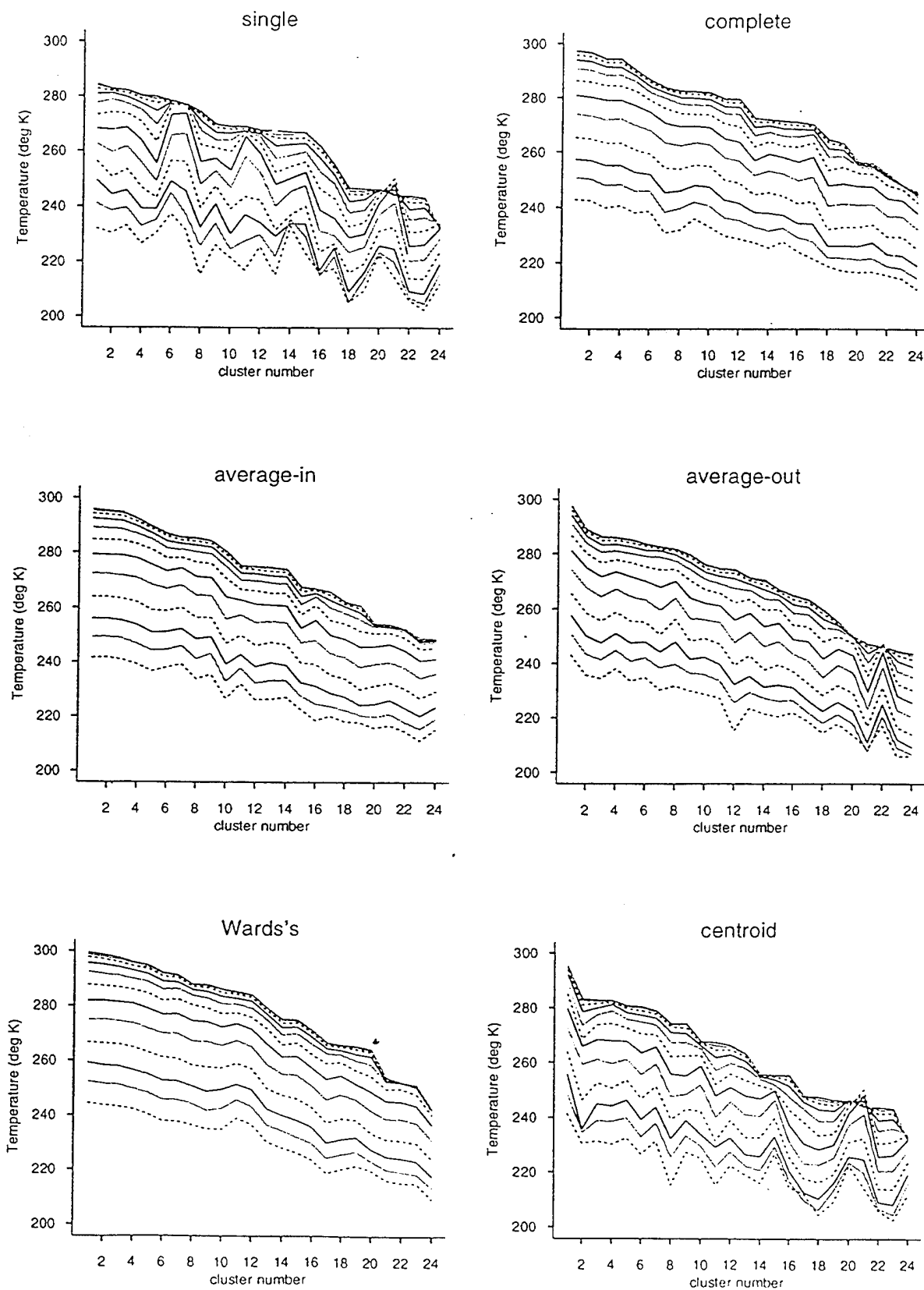
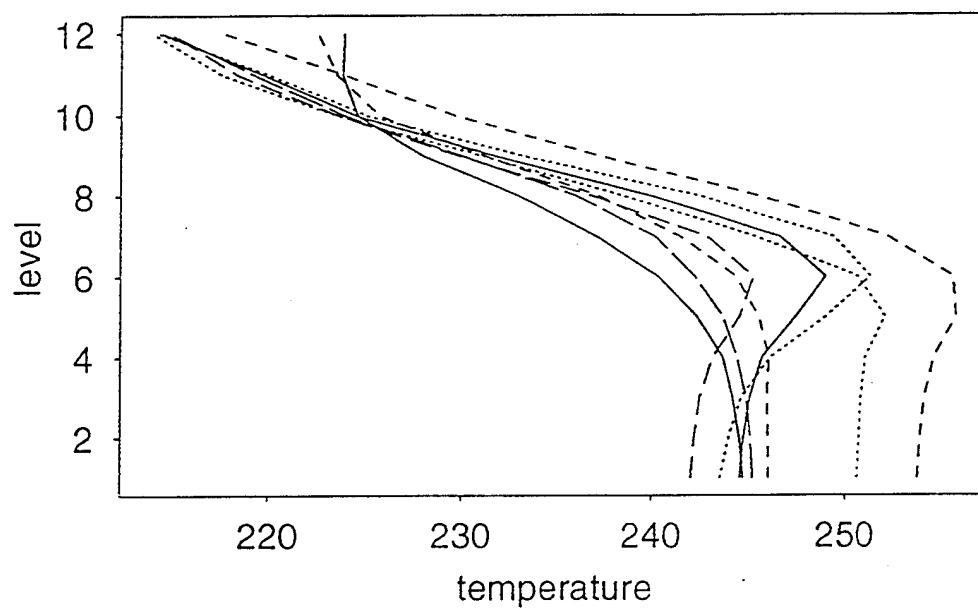


FIGURE 6

## Cluster number 22



**FIGURE 7**

# Precipitable Water

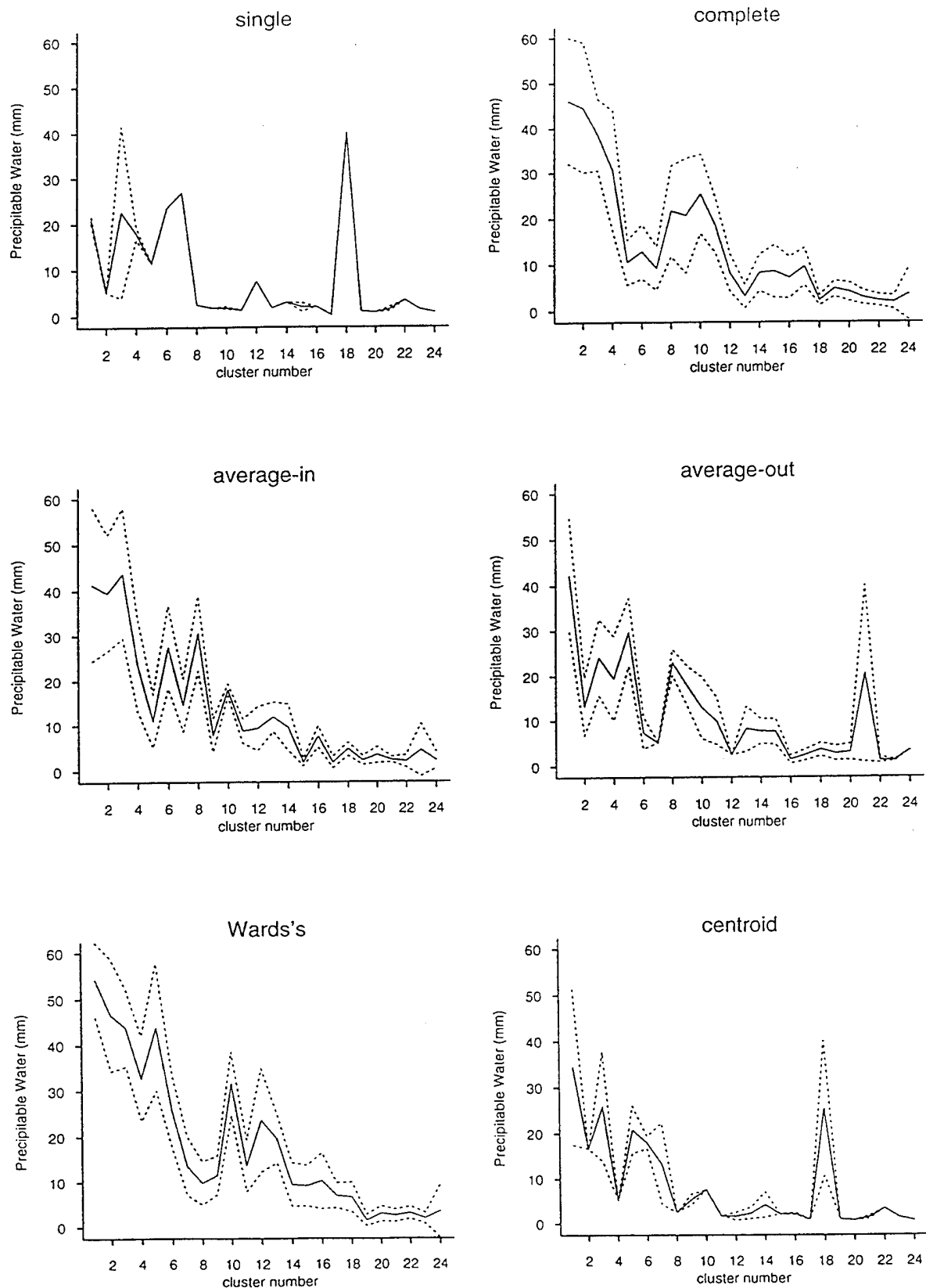


FIGURE 8

## **Appendix B: Vertical stacking of layer cloud amounts**

Total cloud cover, which is the most reliable quantity of the RTNEPH data base, is related to the layer cloud amounts predicted by the various cloud schemes. The exact relation depends on the amount of overlap between the layer clouds. We derived a stacking scheme for the regime-based forecasts by investigating what kind of overlap assumption results in the best agreement between the observed and computed (from layer cloud amounts) total cloud cover in the 1/2-mesh RTNEPH data.

The two limiting cases are random overlap (location of layer clouds completely uncorrelated, resulting in largest total cloud amount), and maximum overlap (location of layer clouds perfectly correlated, resulting in smallest total cloud amount). The formulas relating total cloud cover ( $c_{tot}$ ) to the cloud cover of N cloud layers ( $c_{lay}$ ) for these two cases are given by (all cloud amounts normalized to the interval [0,1]):

$$\text{Random overlap: } c_{tot} = 1 - \prod_{k=1}^N (1 - c_{lay}(k))$$

$$\text{Maximum overlap: } c_{tot} = \max(c_{lay}(k)) , k = 1, N$$

If only two layers are combined, these two formulas can be combined into (see p. A1-A4 of (Mitchell and Hahn, 1989), hereafter referred to as MH):

$$c_{A\_B} = c_A + (1 - c_A)c_B(1 - r),$$

where  $c_{A\_B}$  represents the total cloud cover of layers A and B, and where it is assumed that  $c_A \geq c_B$ . The parameter  $r$  ( $0 \geq r \geq 1$ ) determines the amount of overlap:  $r=0$  ( $r=1$ ) correspond to random (maximum) overlap (note that MH use the quantity  $R=1-r$  in their formula). Intermediate values of  $r$  correspond to partial correlations of layer clouds (note, however, that the quantity  $r$  is not equal to the correlation coefficient: since cloud amounts are confined to the interval [0,1], the combination of the amounts is highly nonlinear and cannot be simply related to the correlation coefficient).

When combining more than 2 layers, the above 2-layer equation must be applied successively until the total cloud amount has been computed. In the case of the compacted 1/2-mesh RTNEPH data, the 6 layers are compacted in 3 steps (as in MH):

Step 1: combine amounts  $c_1$  and  $c_2$  into  $c_{1\_2}$ ,  $c_5$  and  $c_6$  into  $c_{5\_6}$

Step 2: combine amounts  $c_3$  and  $c_{1\_2}$  into  $c_{1\_3}$ ,  $c_4$  and  $c_{5\_6}$  into  $c_{4\_6}$

Step 3: combine amounts  $c_{1\_3}$  and  $c_{4\_6}$  into  $C_{high}$

The notation  $c_{i\_j}$  denotes the amount derived from compacting layers  $i$  through  $j$ ; the layer indices correspond roughly to the mandatory pressure levels (1-1000 hPa; 2-850 hPa; 3-700 hPa; 4-500 hPa; 5-400 hPa; 6-300 hPa).

MH used different values of  $r$  for the different layer combinations. They are plotted in Figure B.1, which shows the value  $r$  as a function of the distance between the layers, expressed as the absolute value of the natural logarithm of the ratio of the two layer pressures ( $\ln(P^A/P^B)$ ). Also shown are curves of  $r$  from the formula

$$r = \frac{1}{1 + \alpha (\ln(P^A/P^B))^2}$$

for two values of  $\alpha$ . The parameter  $\alpha$  can be expressed in terms of the minimum value of  $r$ ,  $r_{\min}$ , which is the value of  $r$  at the maximum layer separation (layers 1 and 6,  $\ln(1000/300) = 1.204$ ). The upper curve ( $r_{\min} = 0.2$ ) was fitted to the value used by MH for the maximum layer separation. Note that random overlap corresponds to a value of  $r_{\min} = 0$ , and maximum overlap to  $r_{\min} = 1$ .



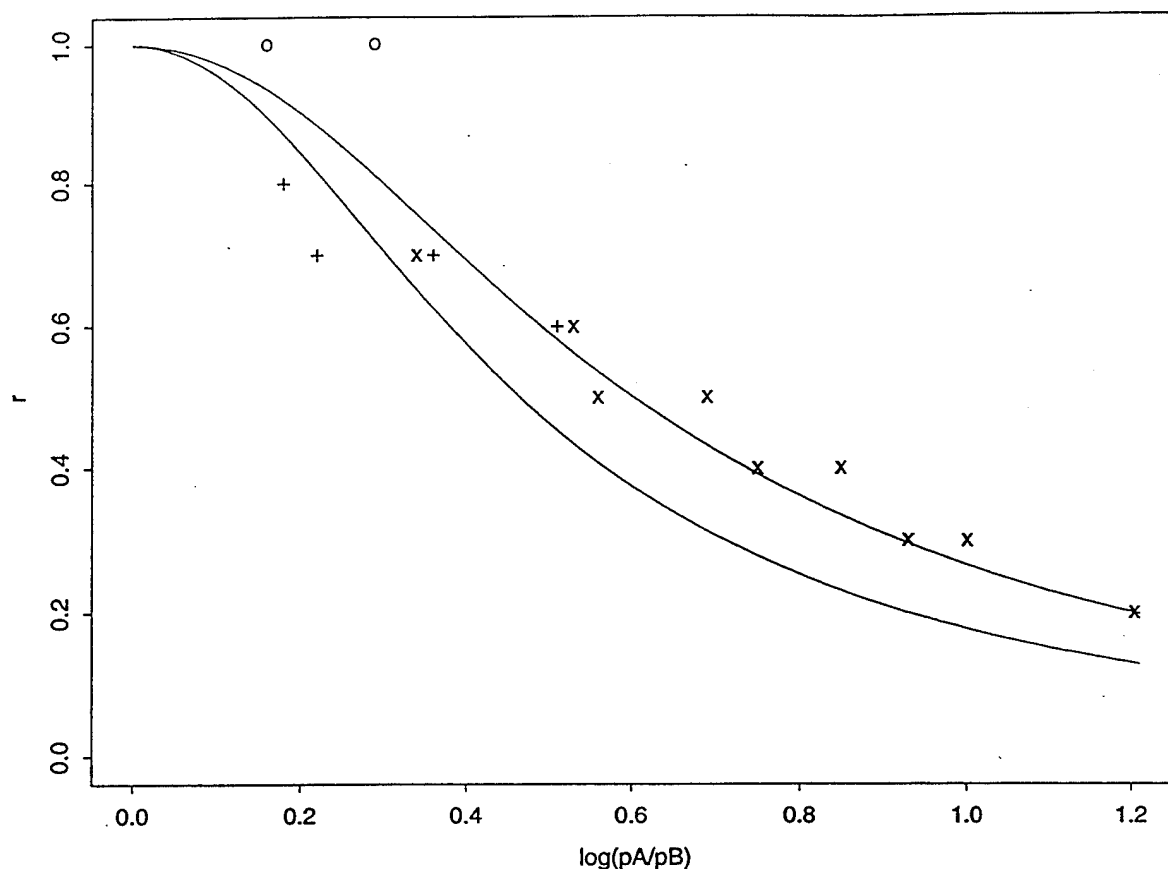


Figure B.1: The overlap parameter  $r$  as a function of the layer separation. Shown are values used by MH during step 1 (circles), step 2 (+), and step 3 (x), as well as from the formula for  $r_{\min} = 0.2$  (top curve) and  $r_{\min} = 0.13$  (bottom curve).

We tested different values of  $r_{\min}$  with the compacted 1/2-mesh data for one time period in January, and the whole month of July. Best agreement between the computed and observed total cloud amounts (using only points with nonzero cloud cover) was found for values of  $r_{\min}$  between 0.12 and 0.14. Based on these results, a value of 0.13 was chosen for  $r_{\min}$  for all computation of total cloud cover from predicted layer amounts for the regime-based schemes<sup>1</sup>. The corresponding curve of  $r$  is plotted as the lower curve in Figure B.1; we note that this corresponds to assuming slightly less overlap between layers than MH.

In the PLScheme, cloud cover is predicted for three terrain-following decks (low, middle, and high) rather than the six pressure levels used in the

<sup>1</sup> Because of a coding error, a value of 0.12 was used in the scheme that used multiple linear regression of the residuals from the regime average cloudiness.

compacted RTNEPH data. Total cloud amount for the PLScheme was computed using the same stacking formula as in (Norquist et al., 1994):

Step 1: combine amounts  $c_{mid}$  and  $c_{low}$  into  $c_{m_l}$

Step 2: combine amounts  $c_{m_l}$  and  $c_{high}$  into  $c_{high}$ .

A value of  $r=.45$  is used in both steps. For the value of  $r_{min}$  used here (0.13), this value of  $r$  corresponds to a layer separation of  $\ln(P^A/P^B) = 0.51$ ; for step 1, this would correspond to a high cloud pressure level of 300 hPa and a middle cloud level pressure of 502 hPa, and for step 2 a low cloud pressure of 850 hPa and a combined middle-high cloud level pressure of 508 hPa. It is thus fairly close to the values used in the regime based schemes for step1, but a slightly larger overlap is assumed in the second step.